

Db 968 -----GlnaCysCysAlaPro---ProAlaPro----- 976
 Qy 315 TCAGAGCCAGTATGCGCATTAATTGCTACAGGGTGTACCGTGCATCCGCTACGC 374
 Db 977 -----CysCysLeuProThrIle 982
 Qy 375 CCAGCCACCCCTGCTGCTGCTGCTACAGTACGAGTACGAGGAGTATGCTGC 434
 Db 983 ProCysCysProIleProCysCysProGln-----ProIleIleCysCys 998
 Qy 435 CGAGCC-----CTACACACACACTGCTCCAGC----- 464
 Db 999 GlnProAlaProValCysLeuProProThrCysSerIleAsnPhelyLeuPro 1018
 Qy 465 -----CCCCACTACGCGCTGTGGTCATGATGCTTTGCG 500
 Db 1019 ThrIleProIleCysGlyArgAlaCysProSerCysProCys 1032
 RESULT 7
 B89716
 protein F45B8.3 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
 C/Accession: B89716
 R:anonymous, The C. elegans Sequencing Consortium.
 A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MIMD:99069613; PMID:9811916
 A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_elegans/
 A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: B89716
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-211 <STO>
 A/Cross-references: GB:chr_X; PIDN:CA805726.1; PID:G3877144; GSPDB:GN00028; CESP:F45B8.3
 C:Genetics:
 A:Gene: F45B8.3
 A:Map position: X
 Alignment Scores:
 Pred. No.: 1.09 Length: 211
 Score: 100.50 Matches: 46
 Percent Similarity: 38.12% Conservative: 15
 Best Local Similarity: 28.75% Mismatches: 65
 Query Match: 8.63% Indels: 34
 Gaps: 6
 US-09-809-545a-1_COPY_535_1143 (1-609) x B89716 (1-211)
 Qy 15 GGCCTGAAACCCCTACACCAATGCGTGAATTAATCCATTGTCGCGCGTCTACAG 74
 Db 56 GlyProGlnProPheAsnProValPheAsnIleHisPheAsnCysCysGlyAlaProArg 75
 Qy 75 CCCCAGCTTATGCG-----AGGACAGGTGCTGTGTGCGCAGGCGCAACGAGAGG 125
 Db 76 ProSerProSerCysCysProTyAlaProProAlaProLeuProProProPro 95
 Qy 126 ATCTTCATGATACAGTGGCCCACTGCTGTATATCTTGTGAAGCTGAGCTTTCC 185
 Db 95 oAlaSerProCysCysGlyProSerProVal-----ProAlaProCysCysPr 111
 Qy 186 ATATCCGCGCCGCTGCGAGTGCATACGAGGCGCTCCTCGAGGCGGTGTCG 245
 Db 111 oProProProAlaProAlaAlaProCys----- 120
 Qy 246 CACCGTGTACAACCTTCAGAGCTGCGCGCCGCCCAATCCGCGCTATGCGCG 305
 Db 121 -----CysProProProProProProProProProProSerProLeuVa 133
 Qy 306 AGTAGTGTATCAAGAGCCAGTGTATGCGCATTAATTGCTACAGGCTGTACGTCGATA 365
 Db 133 lCysCysGlyGlnAlaProValProGluAsnPro-CysCysGlnIleValAlaA-----A 151

Qy 366 CCGCTACGCCCGACCCACCC-----CTGCCACTGCTGCTGCTACAGTACGATTCAGGAGC 422
 Db 151 lAlaIleMetProProProProSerAlaPro-AlaCysCysValAlaAlaProValProThr 170
 Qy 423 AGTTATGCTGCGCGACCCCTACCA-----CCACACTGCTGCTACGCCCC 467
 Db 171 AsnProCysCysGlnProAlaProAlaProArgProAlaProAlaProCysValCysSerAlaPro 188
 RESULT 8
 I50703
 transcription factor GATA-6 - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 26-Aug-1999
 C/Accession: I50703
 R:Lavertiere, A.C.; MacNeill, C.; Mueller, C.; Poelmann, R.E.; Burch, J.B.; Evans, T.
 J. Biol. Chem. 269, 23177-23184, 1994
 A>Title: GATA-4/5/6, a subfamily of three transcription factors transcribed in developing
 A:Reference number: A54720; MIMD:94365018; PMID:8083222
 A:Accession: I50703
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-387 <LAV>
 A/Cross-references: EMBL:U1889; NID:G511483; PIDN:AAA57505.1; PID:G511484
 C/Comment: This transcriptional activator is named for the core, GATA, of the nucleotide
 C:Genetics:
 A:Gene: GATA-6
 A/Superfamily: transcription factor GATA-4; GATA-type zinc finger homology
 C:Keywords: DNA binding; nucleus; transcription factor; transcription regulation; zinc fi
 F:178-231/Domain: GATA-type zinc finger homology <GZFI>
 F:181-205/Region: zinc finger GATA motif
 F:232-285/Domain: GATA-type zinc finger homology <GZFP>
 F:235-259/Region: zinc finger GATA motif
 Alignment Scores:
 Pred. No.: 1.43 Length: 387
 Score: 99.00 Matches: 53
 Percent Similarity: 35.94% Conservative: 16
 Best Local Similarity: 27.60% Mismatches: 50
 Query Match: 8.51% Indels: 73
 Gaps: 11
 US-09-809-545a-1_COPY_535_1143 (1-609) x I50703 (1-387)
 Qy 27 CTACACCAATGCGTGAATTAATCCAGTGTGCGCGCTTACAGCCCGACTTCTA 86
 Db 1 MetTyGlnIleHisPheAlaIleSerAlaSerGlnGly----- 12
 Qy 87 TGCAGGACGCTGCTGTGTCGCGCAGGCGCAACGAGGAGATCTTCATGTAAGTGGCC 146
 Db 13 -----ProAlaProTyAlaSerGlySerProGlyGly--Phe-MethIleSerAlaPr 28
 Qy 147 CAGTTCACTTGTATATCT-----TTCGCAATGCTGCTGCTTCCATA 188
 Db 28 oSerSerProValTyAlaProThrThrArgValGlySerValLeuProThrLeuProTy 48
 Qy 189 T----- 189
 Db 48 rLeuGlnGlyGlyAlaAlaGlnProGlyHisAlaProAlaGlyHisValTrpSerG 68
 Qy 190 -CCGCGCGCACTGCTGCACT-----GCATACGAGGCGCTTACCTTCAGGCGCTGG 242
 Db 68 nProAlaAlaGlnSerProSerTyGlyAlaAlaGlyGlyAlaHisProSerGlyArg-- 87
 Qy 243 TCGACCGGTGTACAACCTTCAGAGCTGCGCGCCGCCCAATC----- 291
 Db 88 -----PheProTySerAlaSerProProValAlaAsnGlyAlaSe 101
 Qy 292 -----CCGCGCTATGCGCGAGTGTATCAAGAGCCAGTATGCG-----ATAAATT 341
 Db 101 rArgGlnGlnTyGlyGlyGlyLeuAlaAlaArgGlnGlnTyGlyAlaLeuProArgPr 121
 Qy 342 GCTACAGGCTGTGATGCTGATACCGCTACGCGCCAGCCCACTTCGCTGCTGCTGCTGC 401

Db 121 oleuAnGlySerTyR-----ProAlaProTyAlaIse 132
Qy 402 CTAC---AGTGACAGTACGAGAGATTATGCTCCGACCCCTACACACACATT-- 456
Db 132 rTyValaIyProGInLeuGlyProAlaIyProAlaIaIaProPheGluAenSerValle 152
Qy 457 -----GCTCCAGCCCC 468
Db 152 uHieCyBeuGInGlyArGAlaIaIaProIlePro 163

RESULT 9
T29707
hypothetical protein ZK180.5 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C/Accession: T29707
R/Author: A. J. Le, T. T.
Submitted to the EMBL Data Library, May 1996
A/Description: The sequence of C. elegans cosmid ZK180.
A/Reference number: Z20669
A/Accession: T29707
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-542 <PAU>
A/Cross-references: EMBL:U58748; PIDN:AAB52969.1; GSPDB:GN00022; CESP:ZK180.5
C/Experimental source: Strain Bristol N2; clone ZK180
C/Genetics:
A/Gene: CESP:ZK180.5
A/Map position: 4
A/Insertions: 3/3; 47/2; 484/3
A/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Alignment Scores:
Pred. No.: 1.41 Length: 542
Score: 99.00 Matches: 51
Percent Similarity: 29.18% Conservative: 17
Best Local Similarity: 21.89% Mismatches: 71
Query Match: 8.51% Indels: 94
DB: 2 Gaps: 8

US-09-809-545A-1_COPY_535_1143 (1-609) x T29707 (1-542)

Qy 106 TGCAGGCCCAACGAGAGGA---TCTTCCATGATACAGTGGCCCACTTCTGATAT 162
Db 29 CyGInProSerGInProAlaCyAlaAenProCyGInGlyAlaSerAlaIaIaTyR 48
Qy 163 ACTTGTCAATGCTGCTTCCATATCCGGCCGACGCTGACGATACGAGGG 222
Db 49 AlaAlaProAlaIaGlyTyRProArGProProIle----- 60
Qy 223 GCTCACCTTCGAGCGCTGCGACCGGTACACACCTTGAGAGCTGCGGCCCA 282
Db 61 -----GlyTyRly-----AlaProProProPro 68
Qy 283 CCCCACATCCGGCTTATGGCGAGTA----- 309
Db 69 ProProProProAlaPheGlyGlyLeuGlyGlyAlaAlaTyRProGlyAlaGlyAla 88
Qy 310 -----GTGTATCAAGAGCAGGTATGGCAAT----- 336
Db 89 GlyGlyGInTyRAspThrGlyValHisGlyGlyAlaGInGlyTyRAlaGlyAlaGIn 108
Qy 337 -----AAATTGCTACAGGTGTTACCTGCGATACCGC 369
Db 109 GlyAlaGInGlyTyRAlaGlyAlaGInGlyAlaGInGlyTyRAlaGlyAlaGIn 128
Qy 370 TACGCCACGCCACCCCTGCTGCT----- 399
Db 129 GlyAlaGInGlyTyRAlaGlyAlaGInGlyAlaGInGlyTyRAlaGlyAla 148
Qy 400 -----GCTTACAGTACGATGACGAGAGGATTTATGCT 432
Db 149 ValGInGlyGlyGInIleAlaSerGInGlyTyRAlaGlyAlaGlyAlaProValSerAla 168

Qy 433 GCCGACCCCTACCAACAC----- 450
Db 169 GlyIySerTyRAnGInGlyProAlaAlaIleAnGlyTyRThrSerGInGIn 188
Qy 451 -----ACCTTGCTCCAGCCCGCACCTPAGCGGTGGCCATGATGCTTTGGCCC 504
Db 189 ValGlyAlaIleAlaProAlaGlyTyRgInGInGlyProAlaAlaValHisValAsp 208
Qy 505 TTGACCGATGCCAAGACTAGGAGCGATGATGATGGTCTGTTCTTTCTTCAATG 564
Db 209 AlaThrAlaIaIaThrSerGluGly-----AlaIleSerGInGIn 221
Qy 565 CAGGCTGATATATACCAAGGGGATACACCGTTTGTCT 603
Db 222 GlyIySerSerAlaAlaGlyGlyTyRgInGInGlyAla 234

RESULT 10
A55726
RNA-binding protein Ews - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: A55726; S45007
R/Plougsattel, B.; Martel, M.G.; Thomas, G.; Delattre, O.
A/Title: Cloning and chromosome localization of the mouse Ews gene.
A/Reference number: A55726; MUID:95130099; PMID:7829090
A/Accession: A55726
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-655 <PDB>
A/Cross-references: GB:K79233; NID:g488512; PIDN:CAA5815.1; PID:g488513
A/Note: authors translated the codon TCA for residue 116 as Thr, GCC for residue 123 as C;Key words: carcinogenesis; nucleus; RNA binding; tandem repeat
F:361-436/Domain: ribonucleoprotein repeat homology <RBM>

Alignment Scores:
Pred. No.: 1.54 Length: 655
Score: 98.50 Matches: 49
Percent Similarity: 34.88% Conservative: 11
Best Local Similarity: 28.49% Mismatches: 71
Query Match: 8.46% Indels: 41
DB: 1 Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x A55726 (1-655)

Qy 85 TATGACGACGCGTGTGTCAGGCC---AACGAGAGGATCTTCCATGTACACT 141
Db 29 TyRAlaGInThrThr-----GInAlaTyRgInGInSerTyRgInTyRly 45
Qy 142 GGGCCGAGTCACTGTATATCTTGTCAATGCTGGCTTCCATTCGGCGCCCACT 201
Db 46 GInProThrAspValSerTyRThrGInAlaGInThrThrAlaThrTyRgInThrAla 65
Qy 202 GCTGACGTGATACGAGGGGCTCACCTTGAGGCGGTGTCGACCGTGTACAAACC 261
Db 66 TyRAlaThrSerTyR-----GlyGInProProThrGlyTyRSerThr 79
Qy 262 TTCAGAGCTGCGGGCCGCCCAATCCGGCTTATGGCGGAGTGTGTATCAAGAG 321
Db 80 -----ProThrAlaProGIn-----AlaTyRSerGIn 88
Qy 322 CCAGTGTATGCAATTAATGTGTACAGGGTGTACGCTGATACCGTACGCCAGCC 381
Db 89 ProValGInGlyTyR-----GlyThrGlyAlaTyRAspSerThrThrAla 103
Qy 382 ACCCTTGCACTGCTGCTTACAGTACAGATGACGAGAGATTATGCTCCGACCC 441
Db 104 ThrValThrThrThrGInAla-----SerTyRAlaIaIaGInSer 116

```

OY      442  TACCACGACACACTTGCTGCAGCCGCCACTTAGCGCGTTGGTGCATGAAGCTTTTCG 501
      117  AAtyrglYthrglnProAlaTyrrProThyrglYgln-----Gln 130
OY      502  CCCTTGACCGATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTGTTCTTCTCA 561
      131  ProthrlAthrAlaProThrrThrrProGlnaspGlyasnLysProAlaGluThrSerGln 150
OY      562  TTGCAGCCTACTATATATACCAAGGGGATACACCGT 597
      151  ProGlnSer-----ThrGlyglYtyrAsnGln 160

RESULT 11
T31328
fibroin - Chinese oak silkworm
C:Species: Antheraea pernyi (Chinese oak silkworm)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31328
R:Sezureau, H.; Tamura, T.; Yukuhiro, K.
Submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, Anth
A:Reference number: Z20995
A:Accession: T31328
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-2639 <SEZ>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C:Genetics:
A:Introns: 14/3

```

Alignment Scores:	
Pred. No.:	1.45
Score:	98.50
Percent Similarity:	41.328
Best Local Similarity:	35.388
Query Match:	8.46%
DB:	2
Length:	2639
Matches:	37
Conservative:	6
Mismatches:	44
Indels:	18
Gaps:	4

US-09-809-545A-1_COPY_535_1143 (1-609) X T31328 (1-2639)

[illegible]

RESULT 12
IS1920
mucin - rhesus macaque (Fragment)
C:Species: Macaca mulatta (rhesus macaque)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: IS1920
R:Am. G.; Luo, G.; Wu, R.
Am. J. Respir. Cell Mol. Biol. 10, 546-551, 1994
#:Title: Expression of MUC2 gene is down-regulated by vitamin A at the transcriptional level

A;Reference number: 151920; MUID:94235322; PMID:8179918
A;Accession: 151920
A;Status: preliminary; translated from GB/EMBL/DBD
A;Molecule type: mRNA
A;Residues: 1-216 <RES>
A;Cross-references: EMBL:U00483; NID:g437054; PIDN:AAA20963.1; PID:g437055
C;Genetics:
A;Gene: MUC2

Alignment Scores:	
Pred. No.:	1,79
Score:	98.00
Percent Similarity:	35.88%
Best Local Similarity:	28.24%
Query Match:	8.42%
DB:	2
Length:	216
Matches:	48
Conservative:	13
Mismatches:	60
Indels:	9
Gaps:	49

US-09-809-545A-1_COPY_535_1143 (1-609) X I51920 (1-216)

[illegible]

nb 28 hrproIlethrthrthrthrthra]athrPro--ThrProthrProthrSerThrg 47

QY 203 CTGCAGCTGCATACCGAGGGGCTCACCCTTCGAGGCCGTGTCGCA--CCGTGTACACACA 25

Db 47 InThrProThr-SerThrProIleThrThrThrThrThrThrAlaThrProThrProThr 66

260 CCTCAGAGCTGCGGCGCCCAACCCCAATCCCGGCTATGGCGAGTAGTGTATCAAG 31

Db 67 ProThrSerThrGlnThr-----ProThrSerThrProIleThr----- 79

320 AGCCAGTGTATGGCAATAAATGCTACAGGGTGGTACGGCTGCATACCGTACGCCAAGC 37

[illegible]

2000

$\frac{1}{2}$

[illegible]

nb 128 SerAenthrrbryaMortbrrPro 135

RESULT 13

hypothetical protein F41E6.11 - *Caenorhabditis elegans*

```
C;Date: 29-Oct-1999 #sequence 29-Oct-1999 #text_change 21-Jan-20
```

R; Sammons, L.; Murray, J.

A;Description: The sequence of *C. elegans* cosmid F41E6.
A;Reference number: 731095

A;Accession: 131880
A;Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-316 <SAM>
A/molecule type: DNA

A; Experimental source: strain Bristol N2; clone F41E6

A;Gene: CESP:F41E6.11

A:Map position: 5
A:Introns: 9/2; 21/3
C:Superfamily: hydroxyproline-rich glycoprotein

Alignment Scores:

Pred. No.:	1.76	Length:	316
Score:	98.00	Matches:	46
Percent Similarity:	30.56%	Conservative:	9
Best Local Similarity:	25.56%	Mismatches:	65
Query Match:	8.42%	Indels:	60
DB:	2	Gaps:	7

US-09-809-545a-1_COPY_535_1143 (1-609) x T31880 (1-316)

```

Qy 52 CCAGTTGTGGCGCGGTCTACAGCCCGACTTCTATGACGAGCAGCGTGTGTGCGCAG 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175 ProValProThProValThProValThGlnProThThrYr----- 185

Qy 112 GCCAACGAGGAGATCTTCATGATGAGTGGCCCGACTTCATGATGATCTTCTGCA 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 -----CysAla 187

Qy 172 ATGCTGCTTTCCATATCCGCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 ValProProCysProAlaProAlaAlaThrProValThThAlaGlnProAlaPro----- 205

Qy 232 CGAGCCGCTGCTGCGACCGCTGTACAAACCTTCAGAGCTGCGCGCGCG-----CCA 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 -----ArgProMetProValThr-----AlaProAlaProAlaCysAlaGln 219

Qy 283 CCCCCAATCCCGCGCTATGCGGAGTAGTGTATCAAGACCGAGTGTAGGCAATAATTG 342
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 ProProCysProAlaThrsnProThrsnProThrsnProThrsnProThrsnProThrsn 325

Qy 343 CTACAGGCTGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 226 ThrsnProAlaThrsnProThrsnProThrsnProThrsnProThrsnProThrsnPro 255

Qy 400 -----GCCCTACAGTGAC 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 GlyThrsnAlaGlnGlyThrsnAlaProAlaThrsnAlaGlnGlyThrsnAlaProAlaPro 275

Qy 412 AGTACGAGCGATTATGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 SerThrsnAlaProAlaThrsnProThrsnProThrsnProThrsnProThrsnProThrsn 295

Qy 469 ACCTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 296 AsnThrsnGlyThrsnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 315
  
```

RESULT 14

A26066 segmentation protein eve - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster
C:Date: 25-Oct-1987 #sequence_revision 25-Oct-1987 #text_change 24-Sep-1999

C:Accession: A26066; B26066; A26636
R:Macdonald, P.M.; Ingham, P.; Struhl, G.

Cell 47, 721-734, 1986
A:Title: Isolation, structure, and expression of even-skipped: a second pair-rule gene

A:Reference number: A26066; MUID:87051744; PMID:2877745

A:Accession: A26066
A:Molecule type: DNA

A:Residues: 1-376 <MAC>
A:Cross-references: GB:M14767; NID:9157386; PIDN:AAA285221; PID:9157387

A:Accession: B26066
A:Molecule type: mRNA

A:Residues: 32-268; 279-376 <MA2>
A:Cross-references: GB:M14767

R:Fraser, M.; Hoey, T.; Rushlow, C.; Doyle, H.; Levine, M.
EMBO J. 6, 749-759, 1987

A:Title: Characterization and localization of the even-skipped protein of Drosophila.
A:Reference number: A26636; MUID:87218536; PMID:2884106

A:Molecule type: mRNA
A:Residues: 1-299, 'L', 301-376 <FRA>
A:Cross-references: GB:X05138; NID:97957; PIDN:CAA28784.1; PID:97958
C:Genetics:
A:Gene: eve (even-skipped)
A:Cross-references: FlyBase:FBgn0000606
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:71-127/Domain: homeobox homology <HOX>

Alignment Scores:

Pred. No.:	1.74	Length:	376
Score:	98.00	Matches:	45
Percent Similarity:	28.65%	Conservative:	8
Best Local Similarity:	24.32%	Mismatches:	56
Query Match:	8.42%	Indels:	76
DB:	2	Gaps:	7

US-09-809-545a-1_COPY_535_1143 (1-609) x A26066 (1-376)

```

Qy 64 GCGGTCTACAGCCCGCGACTTCTATGACGAGCAGTGTGTGCGCAGCCAGCAGAG 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138 AlaValThrsnAspProAlaPheAlaAlaSerIleuGlnAlaAlaAlaAlaAlaAla 157

Qy 124 GGATCTTCCATGATGAGTGGCCCGACTTCATGATGATGATGATGATGATGATGATGAT 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 Gly-----MetProThrsnPro 162

Qy 184 CCATAT-----CGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 ProThrsnAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 177

Qy 241 GGTCCACCGCTGTACAGACCTTCAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 -----ValAlaThrsnProMetMetAlaThrsnMetProMetMetGlyMetProGln 195

Qy 286 -----CCATCCCGCGCTATGCGGAGTAGTGTATCAAGACCGAGTGTATGCG 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 MetProThrsnMetGlyMetProGlyThrsnSerGlyThrsnAlaGlyThrsnProSerProThrsn 215

Qy 334 AATAAATTGCTACAGGAGTGTATGAGCTGATGAGCTGATGAGCTGATGAGCTGATGAGCT 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 Gln-----ThrsnGlyThrsn 220

Qy 394 GCTGCTCTACAGTACGATGACGAGTACGAGAGTATGCTGCGACCGCTGACGACACA 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221 -----ProThrsnThrsnPro 225

Qy 454 CTGCTCCAGCGCCCGCGCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 226 AlaArgProAlaProProThrsnProAlaGly----- 235

Qy 514 GCCAAGACTAGAGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 -----ProHisMetHisHisProHisMetMetGlySerSerAlaThrsnGlySer 251

Qy 574 ATATACCAAGGCGA 588
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 SerThrsnSerAlaGly 256
  
```

RESULT 15

T36581 probable transmembrane protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36581
R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1999
A:Reference number: 221575

A:Accession: T36581
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-811 <OLI>

A;Cross-references: EMBL:AL049826; PIDN:CAB42720.1; GSPDB:GN00070; SCOEDB:SCH24.16c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCH24.16c

Alignment Scores:
Pred. No.: 2.06 Length: 811
Score: 97.00 Matches: 49
Percent Similarity: 35.39% Conservative: 14
Best Local Similarity: 27.53% Mismatches: 63
Query Match: 8.33% Indels: 52
DB: 2 Gaps: 11

US-09-809-545a-1_copy_535_1143 (1-609) x T36581 (1-811)

```
QY 70 TACAGCCCCGACTTGTATGAGCGGCGGCTGTGCTGCAGGCCACAGAGGATCT 129
   :: ||||| ::||| ::||| ::||| ::|||
Db 140 HisgluProaspAlaPheAlaHis-----LeupheArgaspGlnGlnGly 156
   |||
QY 130 TCC-----ATGTACAGTGGCCCGAGTTCATTGTATTAATTGCAATGCTGGCTTT 183
   |||
Db 157 GlyHisProserTyrAspAspGlnAlaAlaValProAlaProAlaProGly--- 175
   |||
QY 184 CCATATCCGCGCCGCACTGCTGCAGCTGCATACCGAGGGGCTCACCTTCGAGGCCGTGCT 243
   |||||
Db 176 ProTyrGlyGlyAlaAlaGlyProGlyGlnTyr----- 186
   |||
QY 244 CGCACCGGTGTACACACCTTGAGAGCTGGGGCGGCCACCCCAATCCGGGCTATGGC 303
   |||
Db 187 -----GlyAlaProProProProProGlyGlnTyrGly 198
   |||
QY 304 GGAGTAGTGTATCAAGAGCCAGTGTATGSCAATAATTGCTACAGGGTGTACGCTGCA 363
   |||
Db 199 GlyHisGlnGlyProAspPro---TyrGlyAsnAlaGlnThrHisGlyGlnTyrGly 217
   |||
QY 364 TACGGCTACGCCAGCCAGCCCGCTGCCACTGCTGCTTACAGTGCAGTTACGAGCA 423
   |||
Db 218 -----SerAlaGlnTyrGlyGlySer----- 224
   |||
QY 424 GTTATGCTGCGGACCCCTACACACACACTTGCTCCAGCCCC---ACCTACGGCGTT 480
   |||||
Db 225 -----AlaAlaThrProGlyGlnHisAla--AlaProAlaProGlyGlnTyrAlaVal 241
   |||
QY 481 GGTCGCAATG-----AATGCTTTTGGCGCCTTGACCGCATGCCAAGACTAGAGCCATGCT 534
   |||
Db 242 AlaProThrProGlyAlaAlaGlyAlaProLeuGlnGlnProGlnProGlyProAlaAla 261
   |||
QY 535 GATGATGTGGGTCTCGTCTTCTTCATTCAGCGCTAGTATATACCAAGGGGGA 588
   |||||
Db 262 Ser-----GlnAlaAlaProGlyGlyGly 270
   |||||
```

Search completed: March 12, 2003, 21:21:36
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 12, 2003, 19:21:08 ; Search time 13.5 Seconds
(without alignments)
3742.086 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 1164
Sequence: 1 atgacataataaaagccgcgt.....acaacgctttgtccatcat 609

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delpop 6.0 , Delpext 7.0	

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB-seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xih
-Q=/cgn2_1/USPTO.spool/US09809545/rnat_11032003_130946_12306/app_query.fasta_1.775
-DB=SwissProt 40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LCOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09809545.qcgn_1.1.10 @rnat_11032003_130946_12306 -NCPU=6 -ICPU=3
-NO_XLXLY -NO_MMAP -HARGOUBERY -NB5_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020.5	87.7	396	1	A2BP_MOUSE
2	1009.5	86.7	397	1	A2BP_HUMAN
3	341.5	29.3	391	1	RAM9_HUMAN
4	116	10.0	3063	1	CA1C_HUMAN
5	115	9.9	3119	1	CA1C_MOUSE
6	103.5	8.9	1170	1	TSP2_BOVIN
7	103	8.8	3124	1	CA1C_CHICK
8	99.5	8.5	2003	1	NTC4_HUMAN
9	99	8.5	387	1	GAT6_CHICK
10	98.5	8.5	376	1	FXE1_HUMAN
11	98.5	8.5	655	1	EMS_MOUSE
12	98	8.4	376	1	HMEV_DROME
13	97	8.3	497	1	BRN1_RAT
14	97	8.3	1943	1	PC15_MOUSE
15	96.5	8.3	1964	1	NTC4_MOUSE
16	96	8.2	766	1	ERYA_DROME
17	95.5	8.2	1172	1	TSP2_HUMAN
18	94	8.1	386	1	HXAD_MOUSE

19	94	8.1	388	1	HXAD_HUMAN
20	93.5	8.0	320	1	CH36_CERCA
21	93	8.0	197	1	MCS_MOUSE
22	93	8.0	495	1	BRN1_MOUSE
23	92.5	7.9	656	1	EMS_HUMAN
24	92	7.9	250	1	RECO_CAUCR
25	92	7.9	500	1	BRN1_HUMAN
26	91	7.8	755	1	RRE1_HUMAN
27	90.5	7.8	450	1	AP2C_HUMAN
28	90.5	7.8	1210	1	PER3_HUMAN
29	90.5	7.8	5179	1	MUC2_HUMAN
30	90	7.7	209	1	CCX1_HUMAN
31	89	7.6	463	1	Y102_MYCTU
32	89	7.6	3149	1	TEGU_EBV
33	88	7.6	401	1	HB9_HUMAN
34	88	7.6	474	1	VTPI_TTVIV
35	88	7.6	585	1	PKNI_MYCTU
36	88	7.5	830	1	SREC_HUMAN
37	87.5	7.5	224	1	Y091_NPVAC
38	87.5	7.5	362	1	NK2C_MOUSE
39	87.5	7.5	640	1	PPCM_HUMAN
40	87	7.4	440	1	GAT4_RAT
41	87	7.5	465	1	FXD1_HUMAN
42	87	7.5	625	1	NIFA_AZOB
43	87	7.5	830	1	SREC_HUMAN
44	87	7.5	920	1	AD19_MOUSE
45	86.5	7.4	261	1	LEG3_RAT

ALIGNMENTS

RESULT 1

A2BP_MOUSE STANDARD; PRT; 396 AA.

AC Q9J143; Q9J145; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ataxin-2-binding protein.
OS A2BP1 OR A2BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Osada N., Kusuda Y., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method."
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE OF 21-396 FROM N.A.
RX MEDLINE=20275619; PubMed=10814712;
RA Shihata H., Huynh D.P., Pulst S.-M.;
RT "A novel protein with RNA-binding motifs interacts with ataxin-2."
RL Hum. Mol. Genet. 9:1303-1313(2000).
CC -!- SUBUNIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC EMBL; AB041596; BA95079.1; -
CC EMBL; AF107204; AAF78292.1; -
CC HSSP; P11940; IGVJ.
DR MGD; MGI:1926224; A2bp1.

```

DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR RNA-binding.
RM DOMAIN 116 192 RNA-BINDING (RRM).
FT CONFLICT 60 G->D (IN REF. 2).
FT CONFLICT 128 R->RFR (IN REF. 2).
FT CONFLICT 369 G->S (IN REF. 2).
SQ SEQUENCE 396 AA; 42590 MW; 9A0C59C5E86739F CRC64;

Alignment Scores:
Pred. No.: 3.48e-76 Length: 396
Score: 1020.50 Matches: 192
Percent Similarity: 96.06% Conservative: 3
Best Local Similarity: 94.58% Mismatches: 7
Query Match: 87.67% Indels: 1
DB: Gaps: 1

US-09-809-545A-1_COPY_535_1143 (1-609) x A2BP_MOUSE (1-396)

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTG 60
DB |||||||
QY 195 MetThrAsnLysLysThrValAsnProTyrThrAsnGlyTyrLysLeuAsnProValVal 214
DB |||||||
QY 61 GAGCGGCTACAGCCCGCACTTCTATGACGAGCAGCGTCTGTGTGCGAGGCAACAG 120
DB |||||||
QY 215 GAlaAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlySerAlaAlaAsnGln 234
DB |||||||
QY 121 GAGGATCTTCATGATGACAGTGGCCCGCACTTCTGTATATACCTTCTGCAATGCTGGC 180
DB |||||||
QY 235 GlnGlySerSerMetCysSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 254
DB |||||||
QY 181 TTTCATATCCGCGCCGCACTGCTGACATACCGAGGGGCTACCTTGAGGCGCT 240
DB |||||||
QY 255 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 274
DB |||||||
QY 241 GGTCCGACCGGTACAAACCTTCAGAGCGGCGCCCGCAACCCCAATCCCGGCTAT 300
DB |||||||
QY 275 GAlaArgThrValTyrAsnThrPheArgAlaAlaAlaProProProLleProAlaTyr 294
DB |||||||
QY 301 GCGCGAGTAGTGTATCAGAGCCAGTGTATGGCAATAAATTCGCTACAGGCTGTTACGCT 360
DB |||||||
QY 295 GAlaGlyValValTyrClnAspGlyPheTyrGlyAlaAsp---lleTyrGlyGlyTyrAla 313
DB |||||||
QY 361 GCATACCGCTACAGCCCGCAACCCCTGCTGCTGCTGCTACAGTACAGTACAGCA 420
DB |||||||
QY 314 AlATyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 333
DB |||||||
QY 421 CGAGTTTATGCTGCGAGCCCTACACCAACACACTGTCTCCAGCCCGCACTACGCGCTT 480
DB |||||||
QY 334 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 353
DB |||||||
QY 481 GGTGCATGAATGCTTTTGGCCCTTACCGATGACCAAGACTAGAGCAATGCTGATGAT 540
DB |||||||
QY 354 GAlaAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgGlyHisAlaAspArg 373
DB |||||||
QY 541 GTGGCTCTGCTTTCTTCTTCATTTGACAGGCTAGTATATACCAAGGGGATACCAAGCTTTT 600
DB |||||||
QY 374 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyValTyrArgAsnArgPhe 393
DB |||||||
QY 601 GGTCCATAT 609
DB |||||||
QY 394 AlaProTyr 396
DB |||||||

RESULT 2
A2BP_HUMAN
ID A2BP_HUMAN STANDARD; PRT; 397 AA.
AC Q9NMB1; Q9NS20;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Ataxin 2-binding protein.
GN A2BP1 OR A2BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA WagaSuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y.,
RA Ninomiya K., Iwayanagi T.;
RA "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN (2)
RP SEQUENCE OF 21-397 FROM N.A.
RX MEDLINE=20275619; PubMed=10814712;
RA Shibata H., Huynh D.P., Pulst S.-M.;
RT "A novel protein with RNA-binding motifs interacts with ataxin-2.";
RL Hum. Mol. Genet. 9:1303-1313(2000).
CC - SUBUNIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
CC - TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MUSCLE AND BRAIN.
CC - SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
-----
DR EMBL; AK001027; BA91472.1; -
DR EMBL; AF107203; AAF78291.1; -
DR HSSP; P11940; ICVJ.
DR MIM; 605104; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW RNA-binding.
FT DOMAIN 117 193 RNA-BINDING (RRM).
FT CONFLICT 92 92 A->T (IN REF. 2).
SQ SEQUENCE 397 AA; 42754 MW; E3E9060B868C79880 CRC64;

Alignment Scores:
Pred. No.: 2.77e-75 Length: 397
Score: 1009.50 Matches: 190
Percent Similarity: 95.57% Conservative: 4
Best Local Similarity: 93.60% Mismatches: 8
Query Match: 86.73% Indels: 1
DB: Gaps: 1

US-09-809-545A-1_COPY_535_1143 (1-609) x A2BP_HUMAN (1-397)

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTG 60
DB |||||||
QY 196 MetThrAsnLysLysThrValAsnProTyrThrAsnGlyTyrLysLeuAsnProValVal 215
DB |||||||
QY 61 GAGCGGCTACAGCCCGCACTTCTATGACGAGCAGCGTCTGTGTGCGAGGCAACAG 120
DB |||||||
QY 216 GAlaAlaValTyrSerProLysPheTyrAlaGlyThrValLeuLeuGlySerAlaAlaAsnGln 235
DB |||||||
QY 121 GAGGATCTTCATGATGACAGTGGCCCGCACTTGTATATACCTTCTGCAATGCTGGC 180
DB |||||||
QY 236 GlnGlySerSerMetCysSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255
DB |||||||
QY 181 TTTCATATCCGCGCCGCACTGCTGACATACCGAGGGGCTACCTTGAGGCGCT 240
DB |||||||

```

D	b		256	PheProTYrProAlaIalathralaIalalaIaTyArgGlyAlaNIleuNhgGlyArg	275
O	y	241	GGTGGCACCCTGGTACACAACCTTCAGAGCTGGCGGCCCCCAATCCGGGCAT	300	
D	b	276	GlyArgThrValTYrAsnThrPheArgAlaIalaproProProIleProAlaTYr	295	
O	y	301	GGCGAGAGTAGTATCAAGAGCCAGTGAATGGCAATAATTGCTACAGGTGGTACGCT	360	
D	b	296	GlyGlyValIalTYrGlAspSelyPheTYrGlyAlaAap--IleTYrGlyGlyTYrAla	314	
O	y	361	GCATACCCGCTACGCCACCCACCCCTGCACCTGGCTGCTGACTTAAGTACGATACGGA	420	
D	b	315	AlaTYrArgTYrAlaGIInProThrProAlaThrAlaIalalySerAspserTYrGly	334	
O	y	421	CGAGTTATGCTGCGCAGCCCTTACCAACACACTTGCTCCAGCCCCCACTCGAGCGTT	480	
D	b	335	ArgValTYrAlaIalAaspProTYrHisHISalaleuAlaProAlaproThrTYrGlyVal	354	
O	y	481	GGTGCATGANTGCTTTTGGCCCTTGAACCATGGCAAGACTAGAGAGCATGCTATAT	540	
D	b	355	GlyAlaMetAlaAlaPhalaProleuthraPalalyThraXerSerHisAlaAasp	374	
O	y	541	GTGGGCTCGGTTCTTTCTTCATTTGACGCTAGTATATACCAAGGGAGATACACGTTT	600	
D	b	375	ValGlyIeuValIeuSerIeuGlnAlaSerIleTYrArgGlyGlyTYrAsnArgPhe	394	
O	y	601	GCTCCATAT 609		
D	b	395	AlaProTYr 397		
R	E	S	RESULT 3		
R	B	M	HUMAN		
I	D		STANDARD;	PRT; 391 AA.	
A	C	043251; GUGW4;			
D	T	16-OCT-2001 (Rel. 40, last created)			
D	T	16-OCT-2001 (Rel. 40, last sequence update)			
D	T	15-JUN-2002 (Rel. 41, last annotation update)			
D	E	Putative RNA-binding protein 9 (RNA binding motif protein 9).			
G	N	RBW9.			
O	S	Homo sapiens (Human).			
O	C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
O	C	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
O	X	NCBI_TaxId=9606;			
R	N	[1]			
R	P	SEQUENCE FROM N.A.			
R	P	TISSUE=Placenta;			
R	A	Collins J.E., Burton J.;			
R	L	Submitted (NOV-1997) to the EMBL/genbank/DBJ databases.			
R	P	[2]			
R	P	SEQUENCE OF 86-391 FROM N.A.			
R	A	Whiteley M.;			
R	L	Submitted (DEC-1999) to the EMBL/genbank/DBJ databases.			
R	C	-1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).			
C	C	-----			
C	C	This SWISS-PROT entry is copyright. It is produced through a collaboration			
C	C	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
C	C	the European Bioinformatics Institute. There are no restrictions on it's			
C	C	use by non-profit institutions as long as its content is in no way com-			
C	C	modified and this statement is not removed. Usage by and for commercial			
C	C	entities requires a license agreement (See http://www.isb-sdb.ch/announce/			
C	C	or send an email to license@isb-sdb.ch).			
C	C	-----			
D	R	EMBL; AL009266; CAA15842.1; ALT_SEQ.			
D	R	EMBL; AL049748; CAB63054.1; --			
D	R	HSSP; P11940; 1CVY			
D	R	Genew; HGNC:9906; RBW9.			
D	R	InterPro; IPR00504; RNA_rec_mot.			
D	R	Pfam; PF00076; xrm; 1.			
D	R	SMART; SMO0360; RRM; 1.			
D	R	PROSITE; PS0102; RRM; 1.			
D	R	PROSITE; PS00030; RRM_RNP_1; 1.			
K	M	RNA-binding.			
F	T	DOMAIN 112 188 RNA-BINDING (RRM).			

SQ SEQUENCE 391 AA; 42284 MW; ED2B76C023D16A7B CRC64;
 Alignment Scores:
 Pred. No.: 1,46e-20 Length: 391
 Score: 341.50 Matches: 93
 Percent Similarity: 48.66% Conservative: 16
 Best Local Similarity: 41.52% Mismatches: 63
 Query Match: 29.34% Indels: 52
 DB: 1 Gaps: 10
 US-09-809-545A-1_COPY_535_1143 (1-609) x RBM9_HUMAN (1-391)
 QY 1 ATGACTATPAAAAAGCCCGGTGAACCCCTACACCAATGGCGTGAATTAATCCAGTTGG 60
 DB 191 MettrnAmLbVbSmecValThrProTyrAlaAsnGlyTrrPbYbSeuSerProValVal 21
 QY 61 GGCGCGCTCTACAGCCCGCAGCTTCTATGACGAGCAGCGGTGCTGTGGCCAGGCCAAC-- 11
 DB 211 GYAlaValAlTYrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu 23
 QY 118 ---CAGGAGGAGTCTTCATGTACATGTAGTGCGGCCCACTTCTACTT-----GTATATCTTCT 16
 DB 231 GYAAspAlaAlaValAlaValProLeuSerSeryArgGlyGlyIleAsnThrTyrIleProLeu 25
 QY 169 GCAATGCGCTGGCTTCCCATATCCG---GCCGCACTGCGTGGCGAGCGATPACCGAGGGGT 22
 DB 251 IleIleProGlyPheProTyrProThrAlaAlaThrAlaAlaAlaPheArgGlyAla 27
 QY 226 CACCTTTCAGAGCCGTGTGTGCACCGCTGTACAACACCTTCAGAGCTGCGGCGGCCACCC 28
 DB 271 HisLeuAlaArgGlyArgGlyArgThrValTyrGlyAlaValArg---AlaValProProThr 28
 QY 286 CCAATCCCGGCTTATGGCGGAGTGTATGTATCAAGACCAAGTGTATGGCAATA----- 33
 DB 290 AlaIleProAlaTyrProGlyValAlaTyrGlnAspGlyPheTyrGly--AlaAspLeuTyr 30
 QY 338 -----AATGCTACAGG-----GTGGTACGCTGCA 36
 DB 309 rIleGluSerAlaAlaAcYsPheArgSerAsnArgValAspMetGlnProThrAspMetHis 32
 QY 364 TACCGCTACGCCACCCACCCACCCCTCCACTGCTGCTGCT-----ACAGTGAC 41
 DB 329 sSerIleuLeuLeuGlnProGlnProProLeuLeuGlnProLeuGlnProLeuThrValTh 34
 QY 412 AGTTACGAGCAGATTATGTGTGTCGCGACCCCTTACCAACACACACTTGCTCCAGCCCCACC 47
 DB 349 rValMetAlaGlyCysThrGlnProThrProThr----- 36
 QY 472 TACGGCGTTGGTGCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGAGCCAT 53
 DB 361 -----MetProLeuProLeuPro--LeuAlaMetG 37
 QY 532 GCTGATGATGGTGGTCTGCTGTTCTTCTTCATTCGACGAGCTAGTATACCAAGGCGATAC 59
 DB 370 uLeuAlaLeuTrr-----ArgValTyrThrGluValAlaIaTh 38
 QY 592 AACCGTTTG 601
 DB 382 rAlaAspLeu 385
 RESULT 4
 CA1C_HUMAN STANDARD; PRT; 3063 AA.
 AC Q99715; Q99716;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor.
 GN COL12A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NX NCBI_TaxId=9606;

[1] SEQUENCE FROM N.A. (LONG AND SHORT FORM), AND PARTIAL SEQUENCE.
RP MEDLINE=97288521; PubMed=9143499;
RA Gerecke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,
Hudson D.L., Champilaud M.-F., Olsen B.R., Burgess R.E.;
RT "Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(XI)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL9A1) to human
RT chromosome 6q12-q13.";
RL Genomics 41:236-242(1997).
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
CC NONTRIPE-HELICAL SEQUENCES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC ISOFORM CHAINS.
CC -1- TISSUE SPECIFICITY: FOUND IN COLLAGEN I-CONTAINING TISSUES: BOTH
CC SHORT AND LONG ISOFORMS APPEAR IN AMNION, CHORION, SKELETAL
CC MUSCLE, SMALL INTESTINE, AND IN CELL CULTURE OF DERMAL
CC FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT
CC ISOFORM IS FOUND IN LUNG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL
CC CARCINOMA CELL LINE.
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END (BY SIMILARITY).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC SIMILARITY).
CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY THE
CC LONG VARIANT IS A PROTEOGLYCAN.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (PACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 4 WFPA DOMAINS.
CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U73778; AAC51244.1; -;
DR EMBL; U73779; AAD40483.1; -;
DR HSSP; P02751; 1TTF.
DR Genew; HGNC:2188; COL12A1.
DR MIM; 120320; -;
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR002035; WMF_A.
DR Pfam; PF000441; fn3; 18.
DR Pfam; PF00092; wva; 4.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF02210; TSPN; 1.
DR PRINTS; PR00014; FNTPERIII.
DR PRINTS; PR00453; WMFADOMAIN.
DR SMART; SM00060; FN3; 17.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; WVA; 4.
DR PROSITE; PS50234; WMFA; 4.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
FT SIGNAL 1 34 POTENTIAL
FT CHAIN 25 3063 COLLAGEN ALPHA 1(XII) CHAIN.
FT DOMAIN 25 114 FIBRONECTIN TYPE-III 1.
FT DOMAIN 140 316 WMFA 1.

FT DOMAIN 333 426 FIBRONECTIN TYPE-III 2.
FT DOMAIN 440 616 WFPA 2.
FT DOMAIN 630 721 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 812 FIBRONECTIN TYPE-III 4.
FT DOMAIN 813 903 FIBRONECTIN TYPE-III 5.
FT DOMAIN 904 998 FIBRONECTIN TYPE-III 6.
FT DOMAIN 999 1085 FIBRONECTIN TYPE-III 7.
FT DOMAIN 1086 1178 FIBRONECTIN TYPE-III 8.
FT DOMAIN 1199 1371 WFPA 3.
FT DOMAIN 1384 1473 FIBRONECTIN TYPE-III 9.
FT DOMAIN 1474 1564 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1565 1652 FIBRONECTIN TYPE-III 11.
FT DOMAIN 1654 1751 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1752 1842 FIBRONECTIN TYPE-III 13.
FT DOMAIN 1843 1932 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1933 2023 FIBRONECTIN TYPE-III 15.
FT DOMAIN 2024 2114 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2115 2202 FIBRONECTIN TYPE-III 17.
FT DOMAIN 2203 2291 FIBRONECTIN TYPE-III 18.
FT DOMAIN 2293 2496 WFPA 4.
FT DOMAIN 2323 2451 FIBRONECTIN TYPE-III 18.
FT DOMAIN 2451 2746 FIBRONECTIN TYPE-III 18.
FT DOMAIN 2747 2898 FIBRONECTIN TYPE-III 18.
FT DOMAIN 2899 2941 FIBRONECTIN TYPE-III 18.
FT DOMAIN 2942 3044 FIBRONECTIN TYPE-III 18.
FT DOMAIN 3045 3063 FIBRONECTIN TYPE-III 18.
FT BINDING 798 798 FIBRONECTIN TYPE-III 18.
FT BINDING 889 889 FIBRONECTIN TYPE-III 18.
FT BINDING 981 981 FIBRONECTIN TYPE-III 18.
FT SITE 862 864 FIBRONECTIN TYPE-III 18.
FT SITE 2779 2781 FIBRONECTIN TYPE-III 18.
FT SITE 2895 2897 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2944 2944 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2947 2947 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2950 2950 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2959 2959 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2965 2965 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2968 2968 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2971 2971 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2971 2971 FIBRONECTIN TYPE-III 18.
FT MOD_RES 2983 2983 FIBRONECTIN TYPE-III 18.
FT MOD_RES 3000 3000 FIBRONECTIN TYPE-III 18.
FT MOD_RES 3003 3003 FIBRONECTIN TYPE-III 18.
FT MOD_RES 3014 3014 FIBRONECTIN TYPE-III 18.
FT MOD_RES 3023 3023 FIBRONECTIN TYPE-III 18.
FT MOD_RES 3026 3026 FIBRONECTIN TYPE-III 18.
FT MOD_RES 3029 3029 FIBRONECTIN TYPE-III 18.
FT CARBOHYD 700 700 FIBRONECTIN TYPE-III 18.
FT CARBOHYD 1763 1763 FIBRONECTIN TYPE-III 18.
FT CARBOHYD 2206 2206 FIBRONECTIN TYPE-III 18.
FT CARBOHYD 2528 2528 FIBRONECTIN TYPE-III 18.
FT CARBOHYD 2679 2679 FIBRONECTIN TYPE-III 18.
FT VARSPLIC 25 1188 FIBRONECTIN TYPE-III 18.
SQ SEQUENCE 3063 AA; 333189 MW; 75FEA79FAB548293 CRC64;

Alignment Scores:
Pred. No.: 0.0469 Length: 3063
Score: 116.00 Matches: 45
Percent Similarity: 34.08% Conservative: 16
Best Local Similarity: 25.14% Mismatches: 54
Query Match: 9.97% Indels: 64
DB: 1 Gaps: 8

US-09-809-545A-1_COPY_535_1143 (1-609) x CALC_HUMAN (1-3063)

QY 148 AGTCACCTGTAATTAATCTTGCAGAAATGCTGCTTCCATATCGGCGCCGACCTGCTGA 207
DB 2078 AaAaAaVAlleLeuGlnProLeuGlnProAaPthrProtyr---Lys11eThrVal1le 2096
QY 208 GCTGCATCCGA-----GGGGCTCACCTTGCAGAGCGCGTGTGCACCGTG----- 252
DB 2097 ALaValTyTgLuAspGlyAspGlyGlyHisLeuThrGlyAsnGlyAyrGthrValGlyLeu 2116

QY 253 -----TACAACTTACAGCTGCGCG 276
 Db 2117 LeuProGlnAsnIleHisIleSerAspGluTrpYrThrArgPheArgValSerTrp 2136
 QY 277 CCCCACCCCATCCCGCGCTATGCGAGAGTGTCAAGAGCGAGTGAAGGCAAT 336
 Db 2137 AspProSerProSerProValLeuGlyTrpIleValTrpLysProValGlySerAsn 2156
 QY 337 AATGTGCTACAG-----GGTGGTACCGTGCATACCGCTACGCCACCCCT 387
 Db 2157 GluProMetGluAlaPheValGlyGluMetThrSerTrpYrThrLeu----- 2171
 QY 388 GCCACTGCTGCTGCTACAGTACAGTACGAGAGATTATGCTGCCAGCCCTACAC 447
 Db 2171 ----- 2171
 QY 448 CACAACTGCTGCACGCCCGACCGCTGCGCTGCGCATGCTTTGCG----- 501
 Db 2172 HisAsnLeuAsnProSerThrTrpAspVal-----AsnValTrpAlaGlnTrp 2188
 QY 502 -----CCCTTGACCGATGCCAAGACTAGAGCGCATGTGATGTG 543
 Db 2189 AspSerGlyLeuSerValProLeuThrAspGlnGlyThr----- 2202
 QY 544 GGTCTCGTCTTTCTTCATTCAGGCTAGTATATACCAAGGGGATACACCGTTT 600
 Db 2203 ---LeuTrpLeuAsnValThrAspLeuThrTrpGlnIleGlyTrpAspThrPhe 2220
 RESULT 5
 CALC_MOUSE STANDARD; PRT; 3119 AA.
 ID Q60847; P70322;
 DT 15-JUL-1998 (Ref. 36, Created)
 DT 15-JUN-2002 (Ref. 41, Last sequence update)
 DT 15-JUN-2002 (Ref. 41, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor.
 GN COL12A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND XIIA-2).
 RC STRAIN=Swiss Webster, and C57BL/6J; TISSUE=Skin;
 RX MEDLINE=96170761; PubMed=8601036;
 RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
 RA Kania A.M., Reichebeger E., Baur S.T., Karimbux N.Y., Taylor R.W.,
 RA Olsen B.R., Nishimura I.;
 RT "Structural variation of type XII collagen at its carboxyl-terminal NCI domain generated by tissue-specific alternative splicing."
 RT J. Biol. Chem. 274:22053-22059(1999).
 RL [2]
 CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE SURFACE OF THE FIBRILS. AND THE COL2 AND NC3 DOMAINS MAY BE LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
 CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDA OF NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS, XIIA-1 (SHOWN HERE), XIIA-2/ERK, XIIA-1 AND XIIA-2, ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE FORM OF COLLAGEN XII MAY CONTAIN HOMOCRIMERS OR ANY COMBINATION OF THE VARIOUS ISOFORMS.
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHOONDRIUM, SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERIOSTEUM.

CC -1- DEVELOPMENTAL STAGE: THE LONG NC3 XIIA ISOFORMS ARE PREDOMINANT AT EARLY STAGES (ED7 AND 11); AT LATER STAGES OF DEVELOPMENT (ED15 AND 17) THE SHORT NC3 XIIA FORMS BECOME THE MAJOR FORMS. AS THE SHORT NC3 FORMS BECOME THE MAJOR PRODUCT, THE LONG SPLICED VARIANT CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER BIRTH.
 CC THE LONG NC1 ISOFORMS, XIIA-1 AND XIIA-1, PEAK IN 15-DAY OLD EMBRYOS AND DECREASE IN 17-DAY OLD ONES. THE EXPRESSION OF THE SHORT NC1 FORM XIIA-2 REMAINS CONSTANT THROUGHOUT LATE STAGES OF EMBRYONIC DEVELOPMENT (ED15 AND ED17).
 CC -1- PFM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END (BY SIMILARITY).
 CC -1- PFM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY SIMILARITY).
 CC -1- PFM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY ISOFORM XIIA-2 IS A PROTEOGLYCAN.
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.
 CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC
 DR EMBL; U25652; AAA97719.1; ALT_SEQ.
 DR EMBL; U57095; AAB07047.1; -.
 DR HSSP; P02751; IFNA.
 DR MGD; MGI:88448; Col12a1.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR003962; FN III_repeat.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR02035; VFMA.
 DR Pfam; PF00041; Fn3; 18.
 DR Pfam; PF00092; Vwa; 4.
 DR Pfam; PF01391; Collagen; 4.
 DR Pfam; PF02210; TSPN; 1.
 DR PRINTS; PR00014; FNTYPEIII.
 DR PRINTS; PR00453; VFMADOMAIN.
 DR SMART; SM00060; FN3; 16.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 4.
 DR PROSITE; PS50234; VFMA; 4.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
 FT SIGNAL 1 24
 FT CHAIN 25 319
 FT DOMAIN 25 114 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 140 316 VFMA 1.
 FT DOMAIN 332 423 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 444 620 VFMA 2.
 FT DOMAIN 634 725 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 726 816 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 817 907 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 908 1002 FIBRONECTIN TYPE-III 6.
 FT DOMAIN 1003 1089 FIBRONECTIN TYPE-III 7.
 FT DOMAIN 1090 1182 FIBRONECTIN TYPE-III 8.
 FT DOMAIN 1203 1375 FIBRONECTIN TYPE-III 9.
 FT DOMAIN 1388 1474 FIBRONECTIN TYPE-III 10.
 FT DOMAIN 1480 1568 FIBRONECTIN TYPE-III 11.
 FT DOMAIN 1569 1652 FIBRONECTIN TYPE-III 12.
 FT DOMAIN 1659 1757 FIBRONECTIN TYPE-III 13.
 FT DOMAIN 1758 1848 FIBRONECTIN TYPE-III 14.
 FT DOMAIN 1849 1938 FIBRONECTIN TYPE-III 15.
 FT DOMAIN 1939 2029 FIBRONECTIN TYPE-III 16.
 FT DOMAIN 2030 2120 FIBRONECTIN TYPE-III 17.
 FT DOMAIN 2121 2208 FIBRONECTIN TYPE-III 18.
 FT DOMAIN 2209 2297 FIBRONECTIN TYPE-III 18.

```

FT DOMAIN 2329 2501 VFPA 4.
FT DOMAIN 2456 2751 NONHELIICAL REGION (NC3).
FT DOMAIN 2752 2899 TRIPLE-HELICAL REGION (COL2)
FT DOMAIN 2900 2942 WITH 1 IMPERFECTION.
FT DOMAIN 2943 3045 NONHELIICAL REGION (NC2).
FT DOMAIN 2943 3045 TRIPLE-HELICAL REGION (COL1)
FT DOMAIN 3046 3119 WITH 2 IMPERFECTIONS.
FT BINDING 802 802 NONHELIICAL REGION (NC1).
FT BINDING 893 893 TO CHONDROITIN SULFATE (POTENTIAL).
FT BINDING 985 985 TO CHONDROITIN SULFATE (POTENTIAL).
FT SITE 866 868 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2784 2786 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2896 2898 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 2945 2945 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2948 2948 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2951 2951 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2960 2960 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2966 2966 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2969 2969 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2972 2972 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2984 2984 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3001 3001 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3004 3004 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3015 3015 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3024 3024 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3027 3027 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3030 3030 HYDROXYLATION (BY SIMILARITY).
FT DOMAIN 869 872 POLY-THR.
FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2212 2212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2684 2684 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2684 2684 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 25 1190 MISSING (IN ISOFORM XIIB-1 AND ISOFORM
FT VARSPLIC 3062 3064 XIIB-2)
FT VARSPLIC 3065 3119 ERY -> GSG (IN ISOFORM XIIB-2 AND ISOFORM
FT VARSPLIC 3119 AA; 340239 MW; 9B1F999C86AB3251 CRC64;
FT SEQUENCE 3119 AA; 340239 MW; 9B1F999C86AB3251 CRC64;
FT SQ
Alignment Scores:
Pred. No.: 0.0567 Length: 3119
Score: 115.00 Matches: 43
Percent Similarity: 33.52% Conserves: 16
Best Local Similarity: 24.43% Mismatches: 59
Query Match: 9.88% Indels: 58
DB: 1 Gaps: 7
US-09-809-545a-1_copy_535_1143 (1-609) x CA1C_MOUSE (1-3119)
QY 148 ATTCTACTGTATATCTCTGCAAGCGCTTCCATATCGGCGCGCACTGCTGCA 207
DB 2084 AanaaValIleLeuGlnProLeuGlnProAspThrProtyr--LysIleThrValIle 2102
QY 208 GGTGCATACCGA-----GGGCTCACCCTTCGAGCGCGTGTGCAACCGTG----- 252
DB 2103 AaIleTyGlnAspGlyAspGlyGlnHisLeuThGlyAsnGlyThrValGlyLeu 2122
QY 253 -----TACACACCTTCAGAGCTGGCGG 216
DB 2123 LeuProGlnIAsnIleHisIlePheAspGlyTrpThrArgPheArgValSerTrp 2142
QY 277 CCCCCCACTTCCCGGCTTGGCGGAGTAGTGATTAAGACGACAGTATGCGCAAT 336
DB 2143 AspProSerProSerProValLeuGlyTyLysIleValTyLysProValGlySerAsn 2162
QY 337 AATATGCTACAG-----GGTGTACGCTACATACCGCTACCGCCAGCCACCTT 387
DB 2163 GlnProMetGlnIAsnIleValGlyGlnValThrSerTyThrIleu----- 2177
QY 388 GCCACTGTCTGCTGCTACAGTGAAGTACGACGAGTTATGCTGCGACCCCTACAC 447

```

```

Db 2177 ----- 2177
QY 448 CACACACTGTCTGCAAGCCCGCACTACGCGCTGTGCTGCAATGCTTT----- 498
Db 2178 HisAsnLeuAsnProSerThrThrTyAspValSerValTyAlaGlnTyAspSerGly 2197
QY 499 -----GGGCTTTCAGCGGATGCGCAAGCTAGAGCCATGCTGATGATGGTCTGCT 552
Db 2198 LeuSerValProLeuThrAspGlnGlyThr-----LeuTyR 2210
QY 553 CTTCTCTTATGACAGGCTAGTATATACCAAGGGGATACACCGTTT 600
Db 2211 LeuAsnValThrAspLeuLeuThrTyGlnValGlyTrpAspThrPhe 2226
RESULT 6
TSP2_BOVIN STANDARD; PRT; 1170 AA.
ID TSP2_BOVIN
AC 095116; 028180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor (Corticotropin-induced secreted protein)
DE (CISP).
CN THBS2 OR TSP2 OR TSP-2.
OS Bos taurus (Bovine); Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lafeuillade M., Keramidis M., Aguesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-552 FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafeuillade B., Pellerin S., Keramidis M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells."
RL J. Cell. Physiol. 167:164-172(1996).
RN [3]
RP SEQUENCE OF 318-831 FROM N.A.
RC TISSUE=Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta."
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VFPC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, X96540; CAA65385.1; -
CC EMBL, X87620; CAA60952.1; -
CC HSSP, P00740; 1EDM.
CC InterPro, IPR000561; EGF-like.

```


DR	InterPro:	IPR001881;	EGF Ca.
DR	InterPro:	IPR001791;	Laminin_G.
DR	InterPro:	IPR000864;	TSP1.
DR	InterPro:	IPR003128;	TSPN.
DR	InterPro:	IPR001007;	VWF_C.
DR	InterPro:	IPR003367;	tsp_3.
DR	Pfam:	PF00008;	EGF; 2.
DR	Pfam:	PF00090;	tsp_1; 3.
DR	Pfam:	PF00093;	vwc_1.
DR	Pfam:	PF02210;	TSPN; 1.
DR	Pfam:	PF02412;	tsp_3; 9.
DR	SMART:	SM00181;	EGF; 3.
DR	SMART:	SM00209;	TSP1; 3.
DR	SMART:	SM00210;	TSPN; 1.
DR	SMART:	SM00214;	VWC; 1.
DR	PROSITE:	PS00022;	EGF_1;
DR	PROSITE:	PS01186;	EGF_2; 1.
DR	PROSITE:	PS50092;	TSP1; 3.
DR	PROSITE:	PS01208;	VWFC; 1.
KM	GlycoProfile:	Cell adhesion;	Calcium-binding; Heparin-binding; Repeat;
KM	GlycoProfile:	Cell adhesion;	Signal.
FT	SIGNAL	1	18
FT	CHAIN	19	1170
FT	DOMAIN	19	232
FT	DOMAIN	318	375
FT	DOMAIN	379	430
FT	DOMAIN	435	491
FT	DOMAIN	492	546
FT	DOMAIN	547	587
FT	DOMAIN	588	645
FT	DOMAIN	646	690
FT	DOMAIN	723	758
FT	DOMAIN	759	781
FT	DOMAIN	793	817
FT	DOMAIN	818	840
FT	DOMAIN	841	878
FT	DOMAIN	879	914
FT	DOMAIN	915	950
FT	DOMAIN	951	1170
FT	DOMAIN	951	1170
FT	SITE	926	928
FT	DISUPEID	266	266
FT	DISUPEID	270	270
FT	DISUPEID	551	562
FT	DISUPEID	556	572
FT	DISUPEID	575	586
FT	DISUPEID	592	608
FT	DISUPEID	599	617
FT	DISUPEID	620	644
FT	DISUPEID	650	663
FT	DISUPEID	657	676
FT	DISUPEID	678	689
FT	CARBOHYD	151	151
FT	CARBOHYD	316	316
FT	CARBOHYD	330	330
FT	CARBOHYD	435	435
FT	CARBOHYD	582	582
FT	CARBOHYD	708	708
FT	CARBOHYD	936	936
FT	CARBOHYD	1067	1067
FT	CONFLICT	535	535
FT	CONFLICT	748	748
EQ	SEQUENCE	1170 AA;	129862 MW; 9GCF1BF55B89A051 CRC64;

Alignment Scores:	
Pred. No.:	0.479
Score:	103.50
Percent Similarity:	34.12%
Best Local Similarity:	24.17%
Query Match:	6.69%
DB:	1
	1.69%
	Gaps:
	12
	Length:
	1170
	Matches:
	51
	Conservative:
	21
	Mismatches:
	69
	Indels:
	71
	Gaps:
	12

US-09-809-545A-1_COPY_535_1143 (1-609) X TSP2_BOVIN (1-1170)

```

OY      12  AAAGCCGTGAACCCCTTACACACATGGCTGTGAATTAAATCCAGTTGTGGGCCGGCTTA
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      380  GluGlyTrpSerProTrrPAlaGluTrpThcIuCysSerAlaThrcysGlySerGlyThr
OY      72  CAGCCCGACT-----TGTATGACAG  91
           |||||
Db      400  GlnGlnArgGlyArgSerCysAspValThrSerAsnThrCysLeuGlyProSerIleGln
OY      92  GCACGGTCTGT-----TGTGCCAGGCCCAACGAGAGGAT-----  127
           |||||
Db      420  ThrTrgAlaCysSerLeuGlyArgCysAspHisAlaArgIleArgGlnAspGlyGlyTrpSer
OY      128  -----CTTCCACATGTACATGGGCCCCAGTTCACTTGAT  160
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      440  HistTrpSerProTrpSerSerCysSerValThrCysGlyValGlyAsnVal-ThrArgIle  459
OY      161  ATACTTCTCAATGCGCTGGCTT--TCATATCCGGCCGGCCACTGCTGCAGCTGCATACC  217
           |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      459  eArgLeucCysAsnSerProValProGlnMetGlyGlyArgSerCys-----  474
OY      218  GAGGGGCTCACTTCGAGGCCCGTGTGCACCGTGTCAACACCTTCAAGAGCTGC--GG  274
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      475  -LysGlySerGly-----ArgGluThrLysAlaCysGlnGly  486
OY      275  CGCCCCCAACCCCAATCCCGGCCATGCGCGAGAGTATCAAGAGCCGATGATGGCA  334
           |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      486  YProProCysProValAlaSpGlyArgTrp-----SerProTrp--  498
OY      335  ATAAATTGCTACAGAGGTGGTTACCTGCATACAGCCGTACGCCACGCCCTCCGCACTG  394
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      499  -----SerProTrpSerAlaCysThrVal-----Th  507
OY      395  CTGCTGCCTACAGTGCAGATTACGGACGAGTATTATGCTGCCAGCCCTACACACAC--  452
           |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      507  rCyAlaGlyGlyIleArgGluArgTrpThrArgValCysAsnSerProGluProGlnHisGly  527
OY      453  -----ACTGCTCCAGAGCCGCCCACTACGCGCGTGTGGCATGATAAAGCTTTGGCCG  502
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      527  yGlyLysAspCysValGlyGlyAlaLysGluGlnGlnMetCysAsnArgLysSerCysSpr  547
OY      503  CCTGACCGATGCCAAGACTACGAGCCCATGC  533
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      547  oIleAspGlyCysLeuSer---AsnProCys  556

RESULT 7
CALC_CHICK
ID      CAIC_CHICK  STANDARD;  PRT;  3124  AA.
AC      P13944: 004509:
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Collagen alpha 1(XII) chain precursor (Fibrochimerin).
GN      COL12A1.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxID=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=White Leghorn;
RA      MEDLINE=92011862; PubMed=1918137;
RA      Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA      Nishida Y., Obara M., Kimata K.;
RT      "The complete primary structure of type XII collagen shows a chimeric
RT      molecule with reiterated fibronectin type III motifs, von Willebrand
RT      factor A motifs, a domain homologous to a noncollagenous region of
RT      type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT      site."
RT      J. Cell Biol. 115:209-221 (1991).
RP      [2]
RP      SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP      2846-2873.

```

RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP 2846-2873.

RP 2846-2873

RA MEDLINE=90062079; PubMed=2584192;
RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
RT "Type XII collagen. A large multidomain molecule with partial
RT homology to type IX collagen".
RL J. Biol. Chem. 264:19772-19778(1989).
RN [3]
RP SEQUENCE OF 2960-3076 FROM N.A.
RX MEDLINE=87317590; PubMed=3476925;
RA Gordon M.K., Gerecke D.R., Olsen B.R.;
RT "Type XII collagen: distinct extracellular matrix component
RT discovered by cDNA cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
RN [4]
RP SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.
RC TISSUE-EMBRYO;
RX MEDLINE=93042014; PubMed=1420368;
RA Trueb J., Trueb B.;
RT "The two splice variants of collagen XII share a common 5' end.";
RL Blochim. Biophys. Acta 1171:97-98(1992).
RN [5]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95370352; PubMed=7642694;
RA Koch M., Bohmann B., Mathieson M., Hagios C., Trueb B., Chiquet M.;
RT "Large and small splice variants of collagen XII: differential
RT expression and ligand binding.";
RL J. Cell Biol. 130:1005-1014(1995).
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX.
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDa OF
CC NONTRIPLE-HELICAL SEQUENCES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYCAN. THE LARGE
CC ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN
CC THE SMALL.
CC -1- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,
CC LIGAMENTS, PERICHONDRITUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE
CC TISSUES CONTAINING TYPE I COLLAGEN.
CC -1- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS, TWO TRIPLE-
CC HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL
CC DOMAINS (NC1, NC2, AND NC3).
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 4 VMFA DOMAINS.
CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@sib-sib.ch).
CC -----
DR EMBL, D00824; BAA00701.1; -
DR EMBL, X61024; CAA43358.1; -
DR EMBL, M17375; AAA48718.1; -
DR EMBL, J05137; AAA48635.1; -
DR EMBL, X67327; CAA47744.1; -
DR PIR, A28037; A28037.
DR PIR, A34485; A34485.
DR HSSP, P20701; ILFA.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003961; FN_III.

DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR002035; WVF_A.
DR Pfam; PF00041; Fn3; 17.
DR Pfam; PF00092; Wva; 4.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF02210; TSPN; 1.
DR PRINTS; PR00014; FNTPFIII.
DR PRINTS; PR00453; VMFADOMAIN.
DR SMART; SM00060; FN3; 17.
DR SMART; SM00210; TSPV; 1.
DR SMART; SM00327; WVA; 4.
DR PROSITE; PS0234; WVFA; 4.
KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 3124
FT DOMAIN 25 114
FT DOMAIN 139 311
FT DOMAIN 332 425
FT DOMAIN 439 615
FT DOMAIN 629 720
FT DOMAIN 721 811
FT DOMAIN 812 904
FT DOMAIN 905 998
FT DOMAIN 999 1085
FT DOMAIN 1086 1178
FT DOMAIN 1199 1371
FT DOMAIN 1384 1473
FT DOMAIN 1474 1565
FT DOMAIN 1566 1654
FT DOMAIN 1655 1755
FT DOMAIN 1756 1846
FT DOMAIN 1847 1936
FT DOMAIN 1937 2027
FT DOMAIN 2028 2118
FT DOMAIN 2119 2206
FT DOMAIN 2207 2295
FT DOMAIN 2327 2500
FT DOMAIN 2455 2750
FT DOMAIN 2751 2902
FT DOMAIN 2903 2945
FT DOMAIN 2946 3048
FT DOMAIN 3049 3124
FT DOMAIN 3086 3096
FT DOMAIN 3111 3123
FT BINDING 797 797
FT BINDING 890 890
FT BINDING 981 981
FT SITE 2448 2440
FT SITE 2899 2901
FT CAROXYD 32 32
FT CAROXYD 1006 1006
FT CAROXYD 1032 1032
FT CAROXYD 1044 1044
FT CAROXYD 1512 1512
FT CAROXYD 1767 1767
FT CAROXYD 2210 2210
FT CAROXYD 2273 2273
FT CAROXYD 2532 2532
FT CAROXYD 2683 2683
FT VARSPPLIC 25 1188
FT CONFLICT 1258 1258
FT CONFLICT 1264 1264
FT CONFLICT 2759 2759
FT CONFLICT 2803 2803
FT CONFLICT 2977 2977
FT CONFLICT 3075 3076
SQ SEQUENCE 3124 AA; 340578 MW; 094285AF87F346CF CRC64;

Alignment Scores:

Pred. No.:	0.545	Length:	3124
Score:	103.00	Matches:	42
Percent Similarity:	34.13%	Conservative:	15
Best Local Similarity:	25.15%	Mismatches:	46
Query Match:	8.85%	Indels:	64
DB:	1	Gaps:	8

DR HSPSP; P08709; 18F9.
 DR Genew; HGNC:7884; NOTCH4.
 DR MIM; 164951; .
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Anx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR01881; EGF_Ca.
 DR InterPro; IPR01438; EGF_11.
 DR InterPro; IPR00800; Notch.
 DR Pfam; PF00008; EGF; 26.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBLDMD.
 DR PRINTS; PR00011; EGFBLMDIN.
 DR PRINTS; PR00012; ENTPEL.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00001; EGF_like; 15.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 11.
 DR PROSITE; PS00022; EGF_1; 28.
 DR PROSITE; PS01186; EGF_2; 21.
 DR PROSITE; PS01187; EGF_CA; 9.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1171 1212
 FT DOMAIN 1212 1246
 FT DOMAIN 1246 1286
 FT REPEAT 1286 1665
 FT REPEAT 1665 1665

FT REPEAT 1666 1698 ANK 2.
 FT REPEAT 1700 1732 ANK 3.
 FT REPEAT 1733 1765 ANK 4.
 FT REPEAT 1766 1798 ANK 5.
 FT DISULFID 28 41 BY SIMILARITY.
 FT DISULFID 35 51 BY SIMILARITY.
 FT DISULFID 53 62 BY SIMILARITY.
 FT DISULFID 68 80 BY SIMILARITY.
 FT DISULFID 74 103 BY SIMILARITY.
 FT DISULFID 105 114 BY SIMILARITY.
 FT DISULFID 122 133 BY SIMILARITY.
 FT DISULFID 127 143 BY SIMILARITY.
 FT DISULFID 145 154 BY SIMILARITY.
 FT DISULFID 160 171 BY SIMILARITY.
 FT DISULFID 165 180 BY SIMILARITY.
 FT DISULFID 182 191 BY SIMILARITY.
 FT DISULFID 198 211 BY SIMILARITY.
 FT DISULFID 205 220 BY SIMILARITY.
 FT DISULFID 222 231 BY SIMILARITY.
 FT DISULFID 238 249 BY SIMILARITY.
 FT DISULFID 243 262 BY SIMILARITY.
 FT DISULFID 264 273 BY SIMILARITY.
 FT DISULFID 280 291 BY SIMILARITY.
 FT DISULFID 285 300 BY SIMILARITY.
 FT DISULFID 302 311 BY SIMILARITY.
 FT DISULFID 318 332 BY SIMILARITY.
 FT DISULFID 326 341 BY SIMILARITY.
 FT DISULFID 343 352 BY SIMILARITY.

Alignment Scores:

Pred. No.: 1.04 Length: 2003
 Score: 99.50 Matches: 51
 Percent Similarity: 30.77% Conservative: 17
 Best Local Similarity: 23.08% Mismatches: 62
 Query Match: 8.55% Indels: 91
 DB: 1 Gaps: 14

US-09-809-545A-1_COPY_535_1143 (1-609) x NTc4_HUMAN (1-2003)

QY 57 TGTGGCCGCGCTCATACGCCCCGACTTCTA----- 86
 Db 918 CysProProGlyPheGlnGlySerIleuCySGlnAspHisValaenProCySGluserArg 937
 QY 87 ---TGCAGCAGCGTGTGTGTGTCAGAGCCGACGAGGAGATTCCATGTACAGTGG 143
 Db 938 ProCySGlnaenGlyAlaThrCysewetalagInProserGly----- 951
 QY 144 CCCCAGTTCACCTTGATATCTTCGCAATGCGCTTCCATATCGGCGCCGACCTGC 203
 Db 952 -----TyleuCySGlnCyValaProGlyTyraSpGlyGlnaenCyS 965
 QY 204 TGCAGCTGCATACGAGGGGCTCACCTTCGAGGCGGTGTCGACCGTGTACACACCTT 263
 Db 966 Ser-----LysGluLeu 969
 QY 264 CAGAGCTGC-----GGCGCCCCCACCCTCAATCCGGCCTA 299
 Db 970 AspAlaCySGlnSerGlnProCySHisaenHisGlyThrCySthProLysProGlyGly 989
 QY 300 TGGCCGAGTAGTGTATCAAGAGCCAGTGTATGCGCAATAATGCTACAGGGGTGTACGC 359
 Db 990 PheHisCyValaCySProProGlyPheVal-----GlyLeuArg 1002
 QY 360 TGC-----ATACCGCTACGCGCCAGCCACCC----- 386
 Db 1003 CySGlnGlyAspValaAspIleuCySleuAspGlnProCySHisProThrGlyThrAlaAla 1022
 QY 387 TGCACCTGCTGCTGCTACAGTACGACGATTACGAGCAGATTATGCTGCCGACCCCTTACA 446
 Db 1023 CySHis-----SerIleuAlaenAlaPheTyCySGlnCySleuPro 1036
 QY 447 CCACACACTTGCTCCAGCCGCCACCTACGGCGTGTGTGTCGCAATGATGCTTTGGCCCTT 506
 |||||
 |||||

Db 1037 GlyHisThr-----GlyGlnTrpCysGlu-----ValGluIle 1047
 QY 507 GACCGATGCCAAGACTAGACCATGC-----TGATGATGTGGG 545
 Db 1048 AsproCysHisSer-----GlnProCysPheHisGlyThrCysGluAlaThrAlaGly 1066
 QY 546 TCTGCTTTCTTTTCATTCATTCAGGCTAGTATATACCAAGGGGATACAA---CCGTTTTCG 602
 Db 1067 SerProLeuGlyPheIleCys-----HisCysProLysGlyPheGluGlyProThrCys 1084
 QY 603 TCC 605
 Db 1085 Ser 1085
 RESULT 9
 GAT6_CHICK
 ID GAT6_CHICK STANDARD; PRT; 387 AA.
 AC P43693;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor GATA-6 (GATA binding factor-6).
 GN GATA6.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBI_TaxID=9031;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=White Leghorn;
 RX MEDLINE=94365018; PubMed=8083222;
 RA Leveriere A.C., Macneill C., Mueller C., Poelmann R.E.,
 RA Burch J.B.E., Evans T.;
 RT "GATA-4/5/6, a subfamily of three transcription factors transcribed
 in developing heart and gut."
 RL J. Biol. Chem. 269:23177-23184 (1994).
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH, AND IN SMALL
 CC - TISSUE SPECIFICITY: LOWER LEVELS IN LUNG, LIVER, OVARY AND HEART.
 CC - SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
 CC
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U11889; AAA57505.1; -
 CC HSSP; P16769; 1GNF.
 DR TRANSFAC; T02684; -
 DR InterPro; IPR000679; Znf_GATA.
 DR InterPro; IPR001164; hrip_like.
 DR Pfam; PF00320; GATA; 2.
 DR PRINTS; PR00619; GATZNFINGER.
 DR SMART; SM00401; Znf_GATA; 2.
 DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
 DR PROSITE; PS50114; GATA_ZN_FINGER_2; 2.
 KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT ZN_FING 181 205 GATA-TYPE 1.
 FT ZN_FING 235 259 GATA-TYPE 2.
 FT DOMAIN 240 244 POLY-THR.
 SQ SEQUENCE 387 AA; 40249 MW; 6EEF457BB34DB2C0 CRC64;

DB: 1 Gaps: 11
 US-09-809-545a-1_copy_535_1143 (1-609) x GAT6_CHICK (1-387)
 QY 27 CTACACCAATGGCGGAAATTAATCCAGTTGGGGCGGCTACAGCCCGCACTTCTA 86
 Db 1 MethylGlnThrLeuAlaIleSerAlaSerGlnGly----- 12
 QY 87 TGCAGGACGCGTCTTGTTCGCGCAACGAGGAGATCTTCATGACAGTGGCCC 146
 Db 13 -----ProIaProtyrAepGlySerProGlyGly---Phe-MethIleSerAlaPr 28
 QY 147 CAGTTCACTTGATATACT-----TTCGAATGCTGCTTCCATA 188
 Db 28 oSerSerProValIlyrValProThrThrArgValGlySerValIleuProThrIleuProty 48
 QY 189 T----- 189
 Db 48 rLeuGlnGlyGlyAlaAlaGlnProGlyHisAlaProIaGlyHisValIlyrSerGI 68
 QY 190 -CCGCGCCGCACTGCTGCAGCT-----GCATACGAGGGCTCACCCTGAGCGCTGG 242
 Db 68 nProIaIaIaGlySerProSerIyryGlyAlaIaGlyAlaHisProSerGlyArg-- 87
 QY 243 TCGCACCGGTACAAACCTTCAGAGCTCGCGCGGCCCAACCCCAATC----- 291
 Db 88 -----PheProtyrSerAlaSerProProValAlaSerGlyAla 101
 QY 292 -----CCGCGCCATGGCGGAGATGCTATCAAGCCAGCTGATATGCG-----AATAATT 341
 Db 101 rArgGlnGlnIlyrGlyGlyGlyLeuAlaIaArgGlnGlnIlyrAlaIleuProAArgPr 121
 QY 342 GCTACAGGCGTGTACCTGATACCGCTACCGCCAGCCACCCCTCCACTGCTGCTGC 401
 Db 121 oLeuAsnGlySerIy-----ProIaProtyrAlaIse 132
 QY 402 CTAC--AGTACAGTTACGAGCAGGATTATGCTCCGAGCCCTACACACACACTT-- 456
 Db 132 rTyValGlyProGlnIleuGlyProIaIyProIaIaIaProIaGlyHisSerValIe 152
 QY 457 -----GCTCGAGCCGCC 468
 Db 152 uHisCysLeuGlnGlyAlaAlaIaProIlePro 163
 RESULT 10
 FXEL_HUMAN
 ID FXEL_HUMAN STANDARD; PRT; 376 AA.
 AC 000356; O75765;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Forkhead box protein F1 (Forkhead-related protein FKHL15) (Thyroid
 transcription factor 2) (TF-2).
 GN FOXE1 OR FKHL15 OR TF2 OR TF2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=97312696; PubMed=9169137;
 RA Chadwick B.P., Obermayr F., Frieschuf A.-M.;
 RT "FKHL15, a new human member of the forkhead gene family located on
 chromosome 9q22."
 RL Genomics 41:390-396 (1997).
 RN
 RP SEQUENCE FROM N.A.
 RA Macchia P.E., Mattei M.-G., Lapi P., Penzi G., Di Lauro R.;
 RT "Cloning, chromosomal localization and identification of two
 polymorphisms in the human thyroid transcription factor 2 gene
 (TF2)."
 RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP VARIANT BAMFORTH-LAZARUS SYNDROME VAL-65.
 RX MEDLINE=98361170; PubMed=9697705;
 RA Clifton-Blyth R.J., Wentworth J.M., Heinz P., Crisp M.S., John R.,
 RA Lazarus J.H., Ludgate M., Charterjee V.K.;
 RT "Mutation of the gene encoding human TTF-2 associated with thyroid
 RT agenesis, cleft palate and choanal atresia."
 RL Nat. Genet. 19:399-401 (1998).
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. COULD BE INVOLVED IN
 CC THYROID GLAND ORGANOGENESIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- TISSUE SPECIFICITY: DETECTED IN ADULT BRAIN, PLACENTA, LUNG,
 CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS, HEART, COLON, SMALL
 CC INTESTINE TESTIS AND THYMUS. EXPRESSION WAS STRONGEST IN HEART AND
 CC PANCREAS.
 CC -1- DISEASE: DEFECTS IN FOXE1 ARE THE CAUSE OF BAMFORTH-LAZARUS
 CC SYNDROME. A DISEASE ASSOCIATED WITH THYROID AGENESIS, CLEFT PALATE
 CC AND CHOANAL ATRESIA.
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use. By non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U89995; AAC51294.1; -
 CC EMBL: Y13386; CA73816.1; -
 CC HSSP: Q63245; ZHPH.
 CC TRANSFAC: T02782; -
 CC Genew: HGNC:3806; FOXE1.
 CC MIM: 602617; -
 CC MIM: 241850; -
 CC InterPro: IPR001766; TF_Fork_head.
 CC Pfam: PF00250; Fork_head; 1.
 CC PRINTS: PR00053; FORKHEAD.
 CC ProDom: PD000425; TF_Fork_head; 1.
 CC SMART: SM00539; FH; 1.
 CC PROSITE: PS00657; FORK_HEAD_1; 1.
 CC PROSITE: PS00658; FORK_HEAD_2; 1.
 CC PROSITE: PS50039; FORK_HEAD_3; 1.
 CC DNA-binding; Nuclear protein; Transcription regulation;
 CC Disease mutation.
 CC DOMAIN 7 12 POLY-PRO.
 CC FT DNA BIND 52 143 FORK-HEAD.
 CC DOMAIN 164 182 POLY-ALA.
 CC VARIANT 65 65 A->V (IN BAMFORTH-LAZARUS SYNDROME).
 CC FT CONFLICT 178 182 MISSING (IN REF. 2).
 CC FT SEQUENCE 376 AA; 38289 MW; 766534A7331572C CRC64;
 SQ
 Alignment Scores:
 Pred. No.: 1.18 Length: 376
 Score: 98.50 Matches: 45
 Percent Similarity: 36.48% Conservative: 13
 Best Local Similarity: 28.30% Mismatches: 68
 Query Match: 8.46% Indels: 33
 DB: Gaps: 6
 US-09-809-545a-1_COPY_535_1143 (1-609) x FOXE1_HUMAN (1-376)
 QY 34 AATGCGTGAATAATCACTGTGGCGCGCTACAGCCCGCACTTATGAGGC 93
 DB 123 AATYTTTIPALeUuSPProAna1aGUaSPme-----PheGUuSerGly 140
 QY 94 ACGGCTCTTGTGGCCAGCAACAGAGGATCTTCATGATGACGTGCGCCCACTTCA 153
 DB 141 SerPheLeuAlaArgArgLysArgPheLysArgSerAspLeu----- 154
 QY 154 CTTGATATATCTTGCAATGCTGCTTTCATATCCGCGCCCACTGCTGACCTGCA 213

DB 155 -----SerThyTyrProAlaTyrMethIAspAlaAlaAlaAlaAla 170
 QY 214 TACCGAGGAGCTACCTTGAAGCCGTGTGCGACCGGTACACACCTTGAGAGCTCG 273
 DB 171 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 190
 QY 274 GGGCCCCCACTCCCAATCCCGGCTATGGCGAGTGTATCAAGAGCCAGCTATGCG 333
 DB 191 AlaArgProPro-----TyrProGlyAlaValTyr----- 200
 QY 334 AATAATGCTACAGAGGTGTTAGCTGCATACCGCTACCGCCACCCCTCCCACT 393
 DB 201 -----AlaGlyTyrAlaProProSerLeuAlaAlaProProValTyr 215
 QY 394 GCTGCTGCTACAGTACAGTACGACGAGTTCATGCT-----GCCAGCCCTAC 444
 DB 216 TyrProAlaAlaSerProGlyProCysArgValPheGlyLeuValProGluArgProLeu 235
 QY 445 CACCAACACCTTGCTCCAGCCCGCCCACTACGCGCTTGTCCTCATATGCTTTGCG 501
 DB 236 SerProGluLeuGlyProAlaProSer---GlyProGlyGlySerCysAlaAlaPheAla 253
 RESULT 11
 EMS_MOUSE STANDARD; PRT; 655 AA.
 ID EMS_MOUSE
 AC Q61545;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE RNA-binding protein EMS.
 GN EMS OR EMSH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=95130099; PubMed=7829090;
 RA Plougastel B., Mattei M.-G., Thomas G., Delattre O.;
 RT "Cloning and chromosome localization of the mouse Ews gene."
 RL Genomics 23:278-281 (1994).
 CC -1- FUNCTION: MIGHT FUNCTION AS A REPRESSOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
 CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use. By non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X79233; CA55815.1; -
 CC MGD: MGI:99960; Ewsh.
 CC InterPro: IPR000504; RNA_rec_mot.
 CC InterPro: IPR001876; Znf_RanBP.
 CC Pfam: PF00641; zf-RanBP; 1.
 CC SMART: SM00360; RRM; 1.
 CC SMART: SM00547; ZNF_RBZ; 1.
 CC PROSITE: PS50096; IQ; FALSE_NEG.
 CC PROSITE: PS50102; RRM; 1.
 CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 CC PROSITE: PS01358; ZF_RANBP2_1; 1.
 CC PROSITE: PS50199; ZF_RANBP2_2; 1.
 CC Nuclear protein; Zinc-finger; Zinc; RNA-binding; Metal-binding;
 CC Calmodulin-binding; Repeat; Methylation; Phosphorylation;
 CC Transcription regulation; Repressor.

FT DOMAIN 1 285 EAD (GLN/PRO/THR-RICH).
 FT DOMAIN 256 285 IO.
 FT DOMAIN 360 446 RNA-BINDING (RRM).
 FT ZN_FING 517 548 RANBP2-TYPE.
 FT DOMAIN 8 285 31 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 8 16 1.
 FT REPEAT 17 27 2.
 FT REPEAT 28 34 3.
 FT REPEAT 35 42 4.
 FT REPEAT 43 50 5.
 FT REPEAT 51 59 6.
 FT REPEAT 60 68 7.
 FT REPEAT 69 75 8.
 FT REPEAT 76 84 9.
 FT REPEAT 85 91 10.
 FT REPEAT 92 110 11.
 FT REPEAT 111 116 12.
 FT REPEAT 117 125 13.
 FT REPEAT 126 136 14.
 FT REPEAT 157 163 15.
 FT REPEAT 164 170 16.
 FT REPEAT 171 177 17.
 FT REPEAT 178 188 18.
 FT REPEAT 189 193 19.
 FT REPEAT 194 201 20.
 FT REPEAT 202 206 21.
 FT REPEAT 207 212 22.
 FT REPEAT 213 218 23.
 FT REPEAT 219 224 24.
 FT REPEAT 225 230 25.
 FT REPEAT 231 238 26.
 FT REPEAT 239 245 27.
 FT REPEAT 246 252 28.
 FT REPEAT 253 259 29.
 FT REPEAT 260 276 30.
 FT REPEAT 277 285 31.
 FT DOMAIN 300 339 ARG/GLY/PRO-RICH.
 FT DOMAIN 453 512 ARG/GLY/PRO-RICH.
 FT MOD_RES 300 300 ARG/GLY/PRO-RICH.
 FT MOD_RES 302 302 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 304 304 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 309 309 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 314 314 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 317 317 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 321 321 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 454 454 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 463 463 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 470 470 METHYLATION (MONO- OR DI-) (BY SIMILARITY).
 FT MOD_RES 489 489 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 493 493 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 499 499 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 502 502 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 505 505 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 562 562 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 564 564 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 571 571 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 574 574 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 580 580 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 588 588 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 591 591 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 595 595 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 599 599 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 602 602 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 606 606 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 614 614 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 632 632 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 635 635 METHYLATION (DI-) (BY SIMILARITY).
 FT MOD_RES 266 266 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 SQ SEQUENCE 655 AA; 68418 MW; 50735EDB54247D69 CRC64;

Alignment Scores:

Pred. No.: 1.2 Length: 655
 Score: 98.50 Matches: 49
 Percent Similarity: 34.88% Conservative: 11
 Best Local Similarity: 28.49% Mismatches: 71
 Query Match: 8.46% Indels: 41
 DB: 1 Gaps: 9
 US-09-809-545a-1_copy_535_1143 (1-609) x EMS_MOUSE (1-655)
 QY 85 TATGACGACGAGGTGTTGTGTCAGAGCC--AACGAGGAGATGTCATGACAGT 141
 DB 29 TYZALAGLINTHR-----GlnAlaTyrGlyGlnGlnSerTyrGlyThrTyrGly 45
 QY 142 GGCCCACTTCACTTGTATATCTTGTCAATGCGCTTGCATATCCGCGCCACT 201
 DB 46 GlnProthraspAlaSerTyrThrGlnAlaGlnThrThrAlaThrTyrGlyGlnThrAla 65
 QY 202 GCTGACAGCTCATACGAGGGCTCACCTTCGAGCGCTGTGACACCGTGTACACACC 261
 DB 66 TYZALATHRSErTyr-----GlyGlnProthrglyTyrSerThr 79
 QY 262 TTACAGACTGGGGGCCCCCAACCCGAGGCTATGCGGAGTATGATCAAG 321
 DB 80 -----ProthralaProGln-----AlaTyrSerGln 88
 QY 322 CCAAGTATGCAATATATGTCACAGGTGTACCGTCAATACCGTACGCCAGCC 381
 DB 89 ProvalGlnGlyTyr-----GlyThrGlyAlaTyrAspSerThrThrAla 103
 QY 382 ACCCTGCGACCTGCTGCTGCTTACAGTACAGTACGAGATTTATGCTCCGACCC 441
 DB 104 ThrValThrThrThrGlnAla-----SerTyrAlaAlaGlnSer 116
 QY 442 TACACACACACTGTGTCAGCCCCCAGCGCTGTGTCAGATGATGCTTTGCG 501
 DB 117 AlATyrGlyThrGlnProAlaTyrProThrTyrGlyGln-----Gln 130
 QY 502 CCTTTCACCATGACCAAGATGAGGATGATGATGATGATGATGATGATGATGATGAT 561
 DB 131 ProthrAlaThrAlaProthraArgProGlnAspGlyAsnLysProAlaGlnThrSerGln 150
 QY 562 TTGACGAGCTGTATATACCAAGGGGATACACCGT 597
 DB 151 ProLinserser-----ThrGlyGlyTyrAsnGln 160
 RESULT 12
 HMEV_DROME STANDARD; PRT; 376 AA.
 ID HMEV_DROME P07667; Q9V5E6;
 AC P06602; P07667; Q9V5E6;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Segmentation protein even-skipped.
 GN EVE OR CG2328.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87051744; PubMed=2877745;
 RA McDonald P.M., Ingham P., Struhl G.;
 RT "Isolation, structure, and expression of even-skipped: a second pair-
 rule gene of Drosophila containing a homeo box.";
 RL Cell 47:721-734 (1986).
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87218536; PubMed=2884106;
 RA Frasch M., Hoey T., Rushlow C., Doyle H., Levine M.;
 RT "Characterization and localization of the even-skipped protein of
 Drosophila.";
 RL EMBO J. 6:749-759 (1987).

[3]
RN SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.F.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Peiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abilaj J.F., Agbayani A., An H.-J., Andrews-Pianko C., Baldwin D.,
RA Ballew R.M., Basay A., Bakendahl J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bandhart D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotfier P.,
RA Butts K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Hsaii M., Kallush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshell A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
RA Palazzolo M., Peterson G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Stimpson M., Skipski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RN "The genome sequence of *Drosophila melanogaster*";
RT Science 287:2185-2195(2000).
RL [4]
RP SEQUENCE OF 1-58 FROM N.A.
RP STRAIN=OR-RC, WA-F, ZMS6, AF-S, and FL-S;
RC MEDLINE=96038621; PubMed=8524036;
RA Ludwig M.Z., Kreitman M.;
RT "Evolutionary dynamics of the enhancer region of even-skipped in
RT *Drosophila*.";
RL Mol. Biol. Evol. 12:1002-1011(1995).
RN [5]
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 70-189.
RP MEDLINE=96134926; PubMed=8557047;
RX Hirsch J.A., Aggarwal A.K.;
RT "Structure of the even-skipped homeodomain complexed to AT-rich DNA:
RT new perspectives on homeodomain specificity.";
RL EMBD J. 14:6280-6291(1995).
CC -1- FUNCTION: MAY PLAY A ROLE IN DETERMINING NEURONAL IDENTITY. MAY BE
CC DIRECTLY INVOLVED IN SPECIFYING IDENTITY OF INDIVIDUAL NEURONS.
CC PAIR-RULE PROTEIN REQUIRED FOR SEGMENTATION; INVOLVED IN
CC TRANSFORMING THE BROAD, SPATIAL, APERIODIC EXPRESSION PATTERNS OF
CC THE GAP GENES INTO A SYSTEM OF PRECISE PERIODIC EXPRESSION
CC PATTERNS OF THE PAIR-RULE AND SEGMENTARY POLARITY GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEOBOX FAMILY.
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC	or send an email to license@ibb.ch
DR	EMBL, M14767, AAA28522.1, -
DR	EMBL, X05138, CAA28784.1, -
DR	EMBL, X05138, CAA28785.1, ALT SEQ.
DR	EMBL, AE003831, AAF58865.1, -
DR	EMBL, U32087, AAB05358.1, -
DR	EMBL, U32088, AAB05359.1, -
DR	EMBL, U32089, AAB05360.1, -
DR	EMBL, U32090, AAB05361.1, -
DR	EMBL, U32091, AAB05362.1, -
DR	PIR, A26065, A26066.
DR	HSSP, P14653, 1B72.
DR	TRANSFAC, T00272, -
DR	FLYbase, FBgn0000606, eve.
DR	InterPro, IPR000047, HTH_repressr.
DR	InterPro, IPR001356, Homeobox.
DR	PFam, PF00046, homeobox, 1.
DR	PRINTS, PR00024, HOMEBOX.
DR	PRINTS, PR00031, HTHREPRESSR.
DR	ProDom, PD000010, Homeobox, 1.
DR	SMART, SMO0389, HOX, 1.
DR	PROSITE, PS00027, HOMEBOX_1, 1.
DR	PROSITE, PS00071, HOMEBOX_2, 1.
DR	DNA-binding; Developmental protein; Embryo, Pair-rule protein;
KW	Transcription regulation; Homeobox; Nuclear protein.
FT	DNA BIND 70 129 HOMEBOX.
FT	DOMAIN 165 179 ALA-RICH.
FT	CONFLICT 300 300 L->V (IN REF. 1).
SO	SEQUENCE 376 AA; 39970 MW; 59058E0634B1BED0 CRC64;

[illegible]

Qy	514	GC	CAAGACTAGGAGCATCATGATGATGGGTCGTCGTCCTTCATTCATGCAGGCTAGT	573
Dy	236	-----	ProhibitinshsProhibitinMeuMetClySerSerAlaThrGlySer	251
Qy	574	ATATACCAAGGGGCA	588	
Dy	252	SerTyrSerAlaGly	256	
RESULT 13				
ID	BRN1_RAT	STANDARD:	PRT;	497 AA.
AC	O63262;			
DT	15-JUL-1998	(Rel. 36, Created)		
DT	15-JUL-1998	(Rel. 38, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Brain-specific Homeobox/POU domain protein 1 (BRN-1 protein).			
GN	POU3F3 OR BRN1 OR BRN-1 OR RH52.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxId=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96070400; PubMed=9405434;			
RA	Schreiber J., Enderich J., Sock E., Schmidt C., Richter-Jansberg C.,			
RA	Wegner M.;			
RT	"Redundancy of class III POU proteins in the oligodendrocyte			
RT	lineage.";			
RL	J. Biol. Chem. 272:32286-32293(1997).			
RN	[2]			
RP	SEQUENCE OF 325-449 FROM N.A.			
RC	TISSUE=Hypothalamus;			
RX	MEDLINE=92228769; PubMed=1348858;			
RA	le Moine C., Young W.S.;			
RT	"RH52, a POU domain-containing gene, and its expression in developing			
RT	and adult rat.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:3285-3289(1992).			
CC	-1- SUBCELLULAR LOCATION: Nuclear (By similarity).			
CC	-1- TISSUE SPECIFICITY: BRAIN.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED FROM EMBRYONIC DAY 11.5 INTO			
CC	ADULTHOOD.			
CC	-1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.			
CC	CLASS-3 SUPERFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use. By non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AJ001641; CAA04893.1; -.			
DR	EMBL; M84644; AAA42041.1; -.			
DR	HSSP; P14859; 1OCT.			
DR	InterPro; IPR001356; Homeobox.			
DR	InterPro; IPR000327; POU domain.			
DR	Pfam; PF00046; homeobox_1.			
DR	Pfam; PF00157; pou_1.			
DR	PRINTS; PR00028; POU_DOMAIN.			
DR	ProDom; PD000010; Homeobox_1.			
DR	ProDom; PD000583; POU domain; 1.			
DR	SMART; SMO0389; HOX; 1.			
DR	SMART; SMO0352; POU; 1.			
DR	PROSITE; PS00027; HOMEBOX_1; 1.			
DR	PROSITE; PS00071; HOMEBOX_2; 1.			
DR	PROSITE; PS00035; POU_1; 1.			
DR	PROSITE; PS00465; POU_2; 1.			
KW	Nuclear protein; DNA-binding; Homeobox.			
FT	DOMAIN 28 49			
FT	DOMAIN 103 114			
FT	DOMAIN 135 143			
FT	POLY-GLY.			
FT	POLY-ALA.			
FT	POLY-PRO.			

[illegible]

PT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Protocadherin 15 precursor.	
GN	PCDH15.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
OX	NCBI_TaxId=10090;	
RP	[1]	
RP	SEQUENCE FROM N.A. PubMed=1138007;	
RX	MEUline=20578757; PubMed=1138007;	
RA	Allegriam K.N., Murcia C.T., Kwon H.Y., Pawlowski K.S., Wright C.G.,	
RA	Woychik R.P.;	
RT	"The mouse Ames waltzer hearing-loss mutant is caused by mutation of	
RL	Pcdh15, a novel protocadherin gene.";	
RL	Nat. Genet. 27:98-102(2001).	
RN	[2]	
RP	TISSUE SPECIFICITY.	
RX	MEUline=21322691; PubMed=11429292;	
RA	Murcia C.T., Woychik R.P.;	
RT	"Expression of the Pcdh15 in the inner ear, nervous system and various	
RL	epithelia of the developing embryo.";	
RL	Mech. Dev. 105:153-166(2001).	
CC	-1- FUNCTION: Calcium-dependent cell-adhesion protein. Essential for	
CC	maintenance of normal retinal and cochlear function. Required for	
CC	inner ear neuroepithelial cell elaboration.	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).	
CC	-1- TISSUE SPECIFICITY: Expressed in brain and sensory epithelium of	
CC	eye, developing inner ear. Also found in the spleen, developing	
CC	dorsal root ganglion, dorsal aspect of neural tube, floor	
CC	plate and endopneural cells adjacent to the neural canal.	
CC	-1- DEVELOPMENTAL STAGE: Highest level of expression is detected at	
CC	embryonic day 16.	
CC	-1- DISEASE: Defects in PCDH15 are the cause of the Ames waltzer (av)	
CC	phenotype. It is characterized by deafness and a balance disorder,	
CC	associated with the degeneration of inner ear neuroepithelia.	
CC	-1- SIMILARITY: CONTAINS 11 CADHERIN DOMAINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@sib-sib.ch).	
CC	-----	
DR	EMBL; AF281899; AAG53891.1; -	
DR	HSSP; P09803; ISUH.	
DR	MGD; MGI:1891428; Pcdh15.	
DR	InterPro; IPR002126; Cadherin.	
DR	Pfam; PF00028; cadherin.10.	
DR	PRINTS; PR00205; CADHERIN.	
DR	SMART; SM00112; CA; 11.	
DR	PROSITE; PS00332; CADHERIN 1; 4.	
DR	PROSITE; PS50268; CADHERIN 2; 11.	
DR	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;	
KW	Signal; MultiGene Family; Vision.	
FT	SIGNAL	1 26
FT	CHAIN	27 1943
FT	DOMAIN	27 1381
FT	TRANSMEM	1382 1402
FT	DOMAIN	1403 1943
FT	DOMAIN	45 152
FT	DOMAIN	153 270
FT	DOMAIN	283 400
FT	DOMAIN	401 514
FT	DOMAIN	515 621
FT	DOMAIN	622 722
FT	DOMAIN	724 824
FT	DOMAIN	825 931
FT	DOMAIN	932 1040
FT	DOMAIN	1042 1149
FT	DOMAIN	1150 1264
FT	DOMAIN	1437 1448
FT	POLY-PRO.	

FT	DOMAIN	1772	1778	POLY.-PRO.	(GLCNAC . .)	(POTENTIAL .)			
FT	CARBOHYD	1804	1812	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	57	57	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	102	102	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	206	206	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	424	424	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	564	564	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	667	667	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	729	729	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	773	773	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	826	826	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	856	856	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	1069	1069	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	1089	1089	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
FT	CARBOHYD	1180	1180	N-LINKED	(GLCNAC . .) <th>(POTENTIAL .)</th>	(POTENTIAL .)			
SO	SEQUENCE	1943 AA;	214816 MM;	E3D7DB9F5F738652	CRG64;				
Alignment Scores:									
Pred. No.:	1.66	Length:	1943						
Score:	97.00	Matches:	46						
Percent Similarity:	38.69%	Conservative:	19						
Best Local Similarity:	27.38%	Mismatches:	65						
Query Match:	8.33%	Indels:	38						
DB:	1	Gaps:	9						
US-09-809-545A-1_COPY_535_1143 (1-609) x PC15_MOUSE (1-1943)									
QY	59	TGGGCGGCGG-----TCTACAGCCCGGACTTCTATGACAGCGCGGTGCTTTGGCC	109						
DB	1672	Trperh1sPheSerPheSerThrLeuProThr11eSerArg1a1Valu1GluLeuG1ySer	1691						
QY	110	AGGCCAACACGAGAGGATCTTCCA-----TGTACAGTGGCCCCCAGTTCACCTTGATATA	163						
DB	1692	GlUPro1snVal1ValThrSerPro1laaPcy5ThrLeu-----	1704						
QY	164	CTTCTGCAATGCTCGGCTTCCATATCCGCGCCACTGCTGACGTCGATACGAGGGG	223						
DB	1705	-----GluLeuSerProPro1leuAluArgProAluG1LeuAluSer	1717						
QY	224	CTCACCTTCGAGGCGGTGTCGACCGCTGTATACACACTTCAGAGCTCGCGGC-----	277						
DB	1718	LeuSerSer1ySaArg1uThrProThrCysAlaSerAspThG1uPro1ySaArg1snSer	1737						
QY	278	-----CCCCACCCCAATCC-----CGGCTATGGCGGAGTATGATCAAG	319						
DB	1738	PheG1u1lealAProH1sProProSer11ePheAlaPro1leuProH1sIbPro1leu	1757						
QY	320	AGCCAGTGTATGGCAATTAATTCCTACAGGAGTGTTACGTCGATACCGCTCGCCAGC	379						
DB	1758	ArgProPro1lealAPheThr1ThrPne-----Pro1leuPro1leuSerProPro1sn	1774						
QY	380	CCACCCCTGACCACTGCTGCTGCTCCTACAGTGACGATTCAGGAGAGTTATGCTGCCGAC	439						
DB	1775	ProProProPro1GlnLeu-----ValThrPheSer1leuPro1leSerThr	1789						
QY	440	CCTACACACACACACTTGTCTCCAGCCGCCCACTACGCGGCTTGTCCTGATGATGCTTTG	499						
DB	1790	ProProThrSer1Ser1Ser1leuPro1leuProPro1leuSer1leuProPro-----ProPro	1807						
QY	500	CGCCCTTGACCGATGCCAAGACTA	523						
DB	1808	ArgPro--ProAlaProArg1leu	1814						
RESULT 15									
ID	NTC4_MOUSE	STANDARD:	PRT: 1964 AA.						
AC	P31695; Q62389; Q62390; Q35442; Q9A1W; Q88314; Q88316; Q9R1X0;								
DT	01-JUN-1993 (Rel. 26, Created)								
DT	01-NOV-1997 (Rel. 35, Last sequence update)								
DT	15-JUN-2002 (Rel. 41, Last annotation update)								
DE	Neurogenic locus notch homolog protein 4 precursor (Notch 4)								
GN	NOTCH4 OR INT3 OR INT-3.								

OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92194507; PubMed=1312643;
 RA Robbins J., Blondel B.J., Gallahan D., Callahan R.;
 RT "Mouse mammary tumor gene int-3: a member of the notch gene family
 RT transforms mammary epithelial cells.";
 RL J. Virol. 66:2594-2599(1992).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=97294599; PubMed=9150355;
 RA Gallahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 RT the NOTCH gene family (NOTCH4).";
 RL Oncogene 14:1883-1890(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Testis;
 RX MEDLINE=96281668; PubMed=8681805;
 RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
 RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
 RT cell-specific mammalian Notch gene.";
 RL Development 122:2251-2259(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Bowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Laeky S.,
 RA Loretz C., Schmidt S., Tipton S., Tralcoff R., Zackone K., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 RT region.";
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracisternal type A particle-mediated activation of the Notch4/int3
 RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
 RT mRNAs by retroviral splicing events.";
 RL J. Virol. 73:5166-5171(1999).
 RN [6]
 RP FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 RT Notch4 in embryonic endothelium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
 RN [7]
 RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 RP OF VAL-1463.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeder E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273 (2001).
 RN [8]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released cell-intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC morphogenesis in the developing vascular system.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide

CC bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytic processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
 CC kidney, and at lower levels in the ovary and skeletal muscle. A
 CC very low expression is seen in the brain, intestine, liver and
 CC testis.
 CC -1- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
 CC embryonic development from 9.0 d.p.c.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TGF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- DISEASE: Loss of the extracellular domain causes constitutive
 CC activation of the Notch protein, which leads to hyperproliferation
 CC of glandular epithelial tissues and development of mammary
 CC carcinomas.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M80456; AAB38377.1; -;
 CC EMBL: U43691; AAC52630.1; -;
 CC EMBL: U43691; AAC52631.1; -;
 CC EMBL: AF030001; AAB82004.1; -;
 CC EMBL: AB016771; BAA32281.1; ALT SEQ.
 CC EMBL: AB016772; BAA32283.1; ALT INIT.
 CC EMBL: AB016773; BAA32284.1; ALT INIT.
 CC EMBL: AB016774; BAA32285.1; -;
 CC PIR: A38072; TWMT3.
 CC HSP: P08709; IBF9.
 CC MGD: MGI:107471; Notch4.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001881; EGF-Ca.
 CC InterPro: IPR001438; EGF II.
 CC InterPro: IPR000800; Notch.
 CC Pfam: PF00008; EGF; 27.
 CC Pfam: PF00023; ank; 6.
 CC Pfam: PF00066; notch; 2.
 CC PRINTS: PRO1415; ANKYRIN.
 CC PRINTS: PRO0010; EGFBLDOD.
 CC PRINTS: PRO1452; NOTCH.
 CC SMART: SM00248; ANK; 5.
 CC SMART: SM00179; EGF_CA; 11.
 CC SMART: SM00001; EGF_like; 15.
 CC SMART: SM00004; NL; 2.
 CC PROSITE: PSS0088; ANK_REPEAT; 5.
 CC PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
 CC PROSITE: PSS00010; ASX_HYDROXYL; 11.
 CC PROSITE: PSS00022; EGF_1; 28.
 CC PROSITE: PSS0186; EGF_2; 21.
 CC PROSITE: PSS0187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;

KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 20
 FT CHAIN 21 1964
 FT CHAIN 1411 1964
 FT CHAIN 1428 1964
 FT CHAIN 1463 1964
 FT CHAIN 21 1443
 FT TRANSMEM 1444 1464
 FT TRANSMEM 1465 1964
 FT DOMAIN 21 60
 FT DOMAIN 61 112
 FT DOMAIN 115 152
 FT DOMAIN 153 189
 FT DOMAIN 191 229
 FT DOMAIN 231 271
 FT DOMAIN 273 309
 FT DOMAIN 311 350
 FT DOMAIN 352 388
 FT DOMAIN 389 427
 FT DOMAIN 429 470
 FT DOMAIN 472 508
 FT DOMAIN 510 546
 FT DOMAIN 548 584
 FT DOMAIN 586 622
 FT DOMAIN 623 656
 FT DOMAIN 658 686
 FT DOMAIN 688 724
 FT DOMAIN 726 762
 FT DOMAIN 764 800
 FT DOMAIN 803 839
 FT DOMAIN 841 877
 FT DOMAIN 878 924
 FT DOMAIN 926 962
 FT DOMAIN 964 1000
 FT DOMAIN 1002 1040
 FT DOMAIN 1042 1081
 FT DOMAIN 1083 1122
 FT DOMAIN 1126 1167
 FT DOMAIN 1168 1208
 FT REPEAT 1209 1242
 FT REPEAT 1243 1282
 FT REPEAT 1282 1657
 FT REPEAT 1657 1691
 FT REPEAT 1691 1724
 FT REPEAT 1724 1757
 FT REPEAT 1757 1757

Alignment Scores:
 Pred. No.: 1.83
 Score: 96.50
 Percent Similarity: 31.46%
 Best Local Similarity: 23.94%
 Query Match: 8.29%
 DB: 1

Length: 1964
 Matches: 51
 Conservative: 16
 Mismatches: 71
 Indels: 75
 Gaps: 13

US-09-809-545a-1_copy_535_1143 (1-609) x NTG4_MOUSE (1-1964)

QY 57 TGTGGGCGCGGTCTACAGCCCGACTTCTA----- 86
 DB 914 CysProPProGlyPheGlnGlyLeuGlyValPheAspAspMetAspGlyCysGluProGln 933
 QY 87 ---TGCAGGCGAGGCGTGTGTGCGAGGCGAAGAGGAGATCTTCATGATGAGTGG 143
 DB 934 ProCysValHisGlySerThrCysValProGlnProSerGly----- 947
 QY 144 CCGCACTGACTTGTATGATCTTGCATGCGCTGCGCTTCCATATGCGGCGCGCACTGC 203
 DB 948 -----TyrValCysGlnCysValPheGlyTyrGlnGlnGln----- 960
 QY 204 TGCAGTGCATACCGAGGCGGCGTTCAGGCGGCGGCGTTCAGGCGGCGTTCAGGCGG 263
 DB 961 CysSerLeuValLeuGlnValCysGlnSerGlnProCysHisValHis----- 976

QY 264 CAGAGCTGCGGCG 323
 DB 977 -----GlyThrCysThrSerArgProGlyGlyPheHisCysAlaCysProGly 993
 QY 324 AGTGTATGCAATAATGCTACAGGCGTGTACGCTGC-----ATA 365
 DB 994 PheVal-----GlyLeuArgCysGlnGlyAspValAspGluCys 1006
 QY 366 CCGCTACG 410
 DB 1007 LeuAspArgProCysHisProSerGlyThrAlaAlaCysHis----- 1020
 QY 411 CAGTTACGAGCAGATTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 470
 DB 1021 SerLeuAlaAspAlaPheTyrCysGlnCysLeuProGlyHisThr----- 1035
 QY 471 CTACGCGCGTGTGCG 530
 DB 1036 -----GlyGlnArgCysGlnValGlnMetAspLeuCysGlnSer---GlnPro 1050
 QY 531 TGTGATGATGATGCGGCTCTCGTTCTTTC-----TTCATTGCAAGC 569
 DB 1051 CysSerArgGlyGlySerCysGlnLeuThrThrGlyProProProGlyPheThrCys--- 1069
 QY 570 TAGTATATACCAAGGCGGATACAA---CCGTTTGTCTCC 605
 DB 1070 ---HisCysProGlyGlyPheGlnGlyProThrCysSer 1081

Search completed: March 12, 2003, 21:19:34
 Job time : 22.5 secs


```

RP SEQUENCE Winkelman J.C.;
RA Chen W., "Intcon Organization of the Human HRNP1 Gene.",
RT "Submitted (Nov-1998) to the EMBL/Genbank/DBJ databases."
RL EMBL; AF229507; AAL71904.1; -.
DR EMBL; AF109120; AAL83409.1; JOINED.
DR EMBL; AF109107; AAL83409.1; JOINED.
DR EMBL; AF109108; AAL83409.1; JOINED.
DR EMBL; AF109109; AAL83409.1; JOINED.
DR EMBL; AF109110; AAL83409.1; JOINED.
DR EMBL; AF109111; AAL83409.1; JOINED.
DR EMBL; AF109112; AAL83409.1; JOINED.
DR EMBL; AF109113; AAL83409.1; JOINED.
DR EMBL; AF109114; AAL83409.1; JOINED.
DR EMBL; AF109115; AAL83409.1; JOINED.
DR EMBL; AF109116; AAL83409.1; JOINED.
DR EMBL; AF109117; AAL83409.1; JOINED.
DR EMBL; AF109118; AAL83409.1; JOINED.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF000076; rrm; 1.
DR SMART; SMO0360; RRM; 1.
DR PROSITE; PSS0030; RRM_1; UNKNOWN.1.
SQ SEQUENCE 418 AA; 44793 MW; 4706FCT40F75D01A CRC64;
```

Alignment scores:		
Pred. No.:	4.37e-88	Length: 418
Score:	1065.00	Matches: 189
Percent Similarity:	98.52%	Conservative: 2
Best Local Similarity:	97.54%	Mismatches: 3
Query Match:	91.32%	Indels: 0
DB:	4	Gaps: 0

QY	1	ATGACTAATAAAAGCCGCTGAACCCCTTACACCAATGGCTGGAATTTAAATCCAGTTGTG	60
DB	216	MeCTTrAsnLysLYsThrValAsnProLYThrAsnGlyTrpLYsLeuAsnProValAl	2359
QY	61	GGCGGGCTTACAGGCCCGCACTTCTATACAGGACCGGCGCTGTGTGCGAGGCCAACAG	120
DB	236	GlyAlaValAlTYSerProGlnPheTYrAlaGlyTrpValLeuLYsCYGAlaAlaAsnGln	2555
QY	121	GAGGAGATCTTCATGTATACAGTGGGCCCAAGTTCACATTGTATACTTTCGCAATCCCTGC	180
DB	256	GlnGlySerSerMetLYSerLYSerLYSerLYaProSerSerLeuValLYTrpTrpSerAlaMetProGly	2757
QY	181	TTTCCATATCCGGCCGCCCACTGCTCAGCTGCATACCGAGGGGCTCACTTCGAGCGCT	240
DB	276	PheProLYrProAlaAlaThrLYaAlaAlaAlaAlaLYrGlnLYaAlaLYsLeuAlaGlyAlaG	295
QY	241	GGTGGCACCCGTATACAAACCTTCAAGAGCTGGCGGCCCAACCCCAATCCCGGCTAT	300
DB	296	GlyAlaGlnTrpValTYrAsnThrPheAsnTrpAlaAlaAlaProProProLYrProAlaTYr	315
QY	301	GGCGGAGTATAGTATCAAGAAGCCATGTATATGCGAATAATTGTCAAGGGGTGTAGCT	360
DB	316	GlyGlyValValTYrGlnGlnProValTYrGlyAsnLYsLeuGlnGlnGlyLYrAla	3359
QY	361	GCATACCGCTAGCCGACGAGCCCAACCTCTGAGCTGCTGCTGCTCAGTGAAGATTACGGA	420
DB	336	AlaTYrAlaGlyTYrAlaGlnProThrProAlaThrAlaAlaAlaLYrSerAspSerTYrGly	355
QY	421	CGAGTTTATGCTGCGGAGCCCTTACACACACACACACTTGTCTCAGGCCCCCACTTACGCGCTT	480
DB	356	ArgValAlTYrAlaAlaAspProLYrTrpTrpAlaAlaAlaAlaAlaProAlaProThrTYrGlyVal	3757
QY	481	GGTGCATGAATGCTTTTGGCGCCTTGACCGAGTGCAGAGCTAGAGAGCATGATGAT	540
DB	376	GlyAlaMetAsnAlaPheAlaProLeuThrAlaPheAlaLYrGlnTrpAlaGlnSerTrpAlaAspAsp	395
QY	541	GTGGGTCTGTTCTTTCTTCAATGCAAGGCTAGATATATCCAAAGGGGATACACCGTTTT	600

Db	396	ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAenArgPhe	415
Qy	601	GCTCATAT	609
Db	416	AlaProTyr	418
RESULT 2			
ID	QBR427	PRELIMINARY;	PRT; 417 AA.
AC	QBR427;		
DT	01-JUN-2002	(TReMBLrel. 21, Created)	
DT	01-JUN-2002	(TReMBLrel. 21, Last sequence update)	
DT	01-JUN-2002	(TReMBLrel. 21, Last annotation update)	
DE	Hexaribonucleotide binding protein 1.		
GN	HRNBPL.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euteleia; Rodentia; Scturognathl; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=SKLETAL MUSCLE;		
RA	Chen W., Winkelman J.C.;		
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF191501; ALA83425.1;		
SQ	SEQUENCE 417 AA; 44638 MW; 96144A2BCDA889F CRC64;		

Alignment Scores:	
Pred. No.:	6,636-88
Score:	1061.00
Percent Similarity:	98.52%
Best Local Similarity:	97.55%
Query Match:	91.15%
DB:	11
Gaps:	0
Length:	417
Matches:	198
Conservative:	2
Mismatches:	3
Indels:	0
Gaps:	0

US-09-809-545A-1_COPY_535_1143 (1-609) x QBR4Z7 (1-417)	
QY 1 ATGACTAATAAAAGCGCGTGAACCCCTTACACCAATGGCTGGAATTTAAATCCAGTTGTG	60
Db 215 MetThrAsnValblysthrValAsnProTyrThrAsnGlyIrpbybLeuAsnProValVal	2340
QY 61 GCGCGGGCTCTACAGCCCGGACTTTTATATGACGGACGGTGGCTGTGTGCGCAGCCACACAG	1200
Db 235 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaAsnGln	2540
QY 121 GAGGATATTCACATGATACAGAGGGGCCAGTTCACCTTATATACTTTGGCAATGCCCGC	1800
Db 255 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly	2740
QY 181 TTTCCATATCCGGCGGCCCACTGTGACGTGCATATCCGAGGGGCTACCTTGAGGCCGT	2400
Db 275 PheProTyrProAlaAlaSerAlaAlaAlaValTyrGlnGlyAlaIleValbLeuArgGlyArg	2940
QY 241 GGTGGCACCGGTGTACACACACTTCAAGACTTGGCGGGCCCCACCCCAATCCGGGGCTAT	3000
Db 295 GlyArgThrGlyTyrAsnThrLeuAlaGlyAlaAlaIleProProProIleProAlaTyr	3140
QY 301 GCGCGAGTATGATTCAAAGACCCAGTGTATGGCAATTAATGTCTACAGGGNGTTACGT	3600
Db 315 GlyGlyValValTyrGlnGlnbProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla	3340
QY 361 GCATACCCCTTACGCCAGCCACCCCTGTGCACCTGTCTGTGCTTACAGTGCACGTTACGGA	4200
Db 335 AlaTyrAlaGlyTyrAlaGlnProThrProAlaIleThrIleAlaIleAlaTyrSerSerAspSerTyrGly	3540
QY 421 CGAGTTTATGCTGCCGACCCCTTACACACACACTTGTCTCAGGCCCCCACTTACAGCGCTT	4800
Db 355 ArgValTyrAlaAlaAspProTyrHisIleThrLeuAlaProAlaIleProThrTyrGlyVal	3740
QY 481 GGTGGCAATGATGCTTTGGCGCCCTTACACCGATGCCAATGCCAAGTACGACATGCTGATAT	5400
Db 375 GlyAlaMetCysAlaIlePheAlaProLeuThrAspAlaLysThrAlaGlySerHisAlaAlaAsp	3940

```

QY 541 GTGGCTCGTCTTTCTTCTTCAATGAGGCTAGTATATACCAAGGGGATACACCGCTTT 600
DB 395 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAsnArgphe 414
QY 601 GCTCCATAT 609
DB 415 AlaProTyr 417

RESULT 3
Q8TAE3 PRELIMINARY; PRT; 392 AA.
AC Q8TAE3;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 1 isoform alpha.
GN HENBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelman J.C.;
RT "Molecular Cloning and Chromosomal Localization of a Novel Human
RT Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene
RT Homologous to fox-1 in Caenorhabditis elegans.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelman J.C.;
RT "Molecular Cloning and Chromosomal Localization of a Novel Human
RT Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene
RT Homologous to fox-1 in Caenorhabditis elegans.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RT "The Exon-Intron Organization of the Human HENBP1 Gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DB EMBL; AF094849; AAL83405.1; -.
DR EMBL; AF109120; AAL83408.1; JOINED.
DR EMBL; AF109107; AAL83408.1; JOINED.
DR EMBL; AF109108; AAL83408.1; JOINED.
DR EMBL; AF109109; AAL83408.1; JOINED.
DR EMBL; AF109110; AAL83408.1; JOINED.
DR EMBL; AF109111; AAL83408.1; JOINED.
DR EMBL; AF109112; AAL83408.1; JOINED.
DR EMBL; AF109113; AAL83408.1; JOINED.
DR EMBL; AF109114; AAL83408.1; JOINED.
DR EMBL; AF109115; AAL83408.1; JOINED.
DR EMBL; AF109116; AAL83408.1; JOINED.
DR EMBL; AF109117; AAL83408.1; JOINED.
DR EMBL; AF109118; AAL83408.1; JOINED.
DR EMBL; AF109119; AAL83408.1; JOINED.
SQ SEQUENCE 392 AA; 42084 MW; 68F8CFE2FC6D674 CRC64;

Alignment Scores:
Pred. No.: 2.21e-74 Length: 392
Score: 912.00 Matches: 172
Percent Similarity: 85.71% Conservative: 2
Best Local Similarity: 84.73% Mismatches: 3
Query Match: 78.35% Indels: 26
DB: 4 Gaps: 1

US-09-809-545a-1_COPY_535_1143 (1-609) x Q8TAE3 (1-392)
QY 1 ATGACTAATAAAGAGCCGCTGAAACCCCTACACCAATGGCTGAATTAAATCAGTTGTC 60
DB 216 MetThrAnuLysIleThrValAsnProTyrThrAsnGlyTyrIleValAsnProTyrVal 235
QY 61 GCGCGGCTCTACAGCCCGGACCTTCTATGACGACGCGTGTGTGTCAGGCCCAACG 120
DB 236 GlyAlaValTyrSerProGluPheTyrAlaGlyThrValLeuLeuGlnAlaAsnGln 255

```

```

QY 121 GAGGATCTTCATGATACAGTGGCCCGAGTTCACTGTATATACTTGTCAATGCTTGC 180
DB 256 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 275
QY 181 TTTCATATCCGGCGCCACTGCTGCTGACGTCGATACGAGGGGCTCAGCGGCGT 240
DB 276 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrGlyAlaAlaIleValGlyArg 295
QY 241 GGTGCACCGGTGTACACACCTTCAGAGCTGCGGCGCCCGCCCAATCCCGCTAT 300
DB 296 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProIleProAlaTyr 315
QY 301 GCGGAGTAGTGTATCAAGAGCCAGCTGTATGCAATTAATTCTTACAGGCTGTACGCT 360
DB 316 GlyGlyAlaValTyrGlnGluProValTyrGlyAsnIleLeuGlnGlyGlyTyrAla 335
QY 361 GCATACCGCTTACGCCAGCCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 336 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 355
QY 421 CGAGTTATGCTGCGCAGACCCCTTACACACACTTGTCTCCAGCCCGCCACTTACGCGCT 480
DB 356 ArgValTyrAlaAlaAspProTyrHisAlaAlaAlaProAlaProThrTyrGlyVal 375
QY 481 GGTGCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGAGCCATGTGATGAT 540
DB 376 GlyAlaMet----- 378
QY 541 GTGGCTCGTCTTTCTTCTTCAATGAGGCTAGTATATACCAAGGGGATACACCGCTTT 600
DB 379 -----AlaSerIleTyrArgGlyGlyTyrAsnArgphe 389
QY 601 GCTCCATAT 609
DB 390 AlaProTyr 392

RESULT 4
Q8TAE2 PRELIMINARY; PRT; 395 AA.
AC Q8TAE2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 1 isoform beta.
GN HENBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelman J.C.;
RT "Molecular Cloning and Chromosomal Localization of a Novel Human
RT Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene
RT Homologous to fox-1 in Caenorhabditis elegans.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RT "The Exon-Intron Organization of the Human HENBP1 Gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DB EMBL; AF109106; AAL83406.1; -.
DR EMBL; AF109107; AAL83407.1; JOINED.
DR EMBL; AF109108; AAL83407.1; JOINED.
DR EMBL; AF109109; AAL83407.1; JOINED.
DR EMBL; AF109110; AAL83407.1; JOINED.
DR EMBL; AF109111; AAL83407.1; JOINED.
DR EMBL; AF109112; AAL83407.1; JOINED.
DR EMBL; AF109113; AAL83407.1; JOINED.
DR EMBL; AF109114; AAL83407.1; JOINED.
DR EMBL; AF109115; AAL83407.1; JOINED.

```

DR EMBL; AF109116; AAL83407.1; JOINED.
 DR EMBL; AF109117; AAL83407.1; JOINED.
 DR EMBL; AF109118; AAL83407.1; JOINED.
 DR EMBL; AF109119; AAL83407.1; JOINED.
 SQ SEQUENCE 395 AA; 42402 MW; 2259F9127AA1DE39 CRC64;

Alignment Scores:

Pred. No.:	7,566-66	Length:	395
Score:	818.00	Matches:	158
Percent Similarity:	87.85%	Conservative:	1
Best Local Similarity:	87.29%	Mismatches:	3
Query Match:	70.27%	Indels:	19
DB:	4	Gaps:	1

US-09-809-545a-1_copy_535_1143 (1-609) x 08TAF2 (1-395)

```

QY 1 ATGACTAATAAAGCCGCGTAACCCCTACCAATGGCTGGAATTAATCCAGTTGG 60
DB 216 MettrAenLYbThrValaenProTYrThraenGlyTrpLYbLeuAenProValVal 235
QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGCAGCGTCTGTGCGAGGCCAACAG 120
DB 226 GYAlaValaLYSerProGluPheryrAlaGlyThrValaLeuLeuCYsGlnAlaenGln 255
QY 121 GAGGATCTTCATGATGACAGTGCGCCCACTTCACTTGAATATCTTGCATGCTGCG 180
DB 256 GluGlySerSerMetTYrSerAlaProSerSerLeuValTYrThSerAlaMetProGly 275
QY 181 TTTCATATCCGCGCGCCCACTGCTGAGTGCATACCGAGGGGCTCACTTGAGGCCGT 240
DB 276 PheProTYrProAlaAlaThraAlaAlaAlaTYrAGLYAlaAlaLYbLeuArGlyArG 295
QY 241 GGTCCACCGGTGTACAAACCTTACAGAGCTGGCGGCCCAACCCCAATCCCGGCTAT 300
DB 296 GYAlaGThrValTYrAsnThrPheArGAlaAlaAlaProProProProleProAlaTYr 315
QY 301 GCGGAGTAGTGTATACAGCCAGCTGTATGCAATTAATCTTACAGGTTGTTACGCT 360
DB 316 GYgLYAlaValaTYrGlnGlnProValTYrGlyAenLYbLeuGlnGlnGlyTYrAla 335
QY 361 GCATACCGCTAGCCCGCCAGCCCACTGCTGCTGCTGCTAC----- 405
DB 336 AlaTYrArGTYrAlaGlnProThrProAlaThraAlaAlaTYrSerArpArGlnGln 355
QY 406 -----AGTACAGATTACGAGCAGATT 426
DB 356 PheValaPheValaAlaAlaAspGluLYSerCYsArnThrSer-AlaValaThraAspGluPh 375
QY 427 TATGCTGCCGACCCCTACACACACACTTGTCTCCAGCCCCCACTTACGCGCTGTGCGC 486
DB 375 ewetLeuProThrProThrThrThriSLeuGlnInProProProThraAlaLeuValPr 395
QY 487 A 487
DB 395 o 395

```

RESULT 5

095K10

PRELIMINARY:

PRT: 376 AA.

AC 095K10. 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypoetical 40.6 kDa protein.
 OS Macaca fascicularis (Crib eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Cercopitheciinae; Macaca.
 OC NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TEMPORAL LOBE RIGHT;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,

RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB060859; BAB46877.1; -
 DR InterPro; IPR000504; RNA_rec_not.
 DR Pfam; PF00076; Trm; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 DR Hypoetical protein.
 SQ SEQUENCE 376 AA; 40565 MW; DAA7A4D0D7B1030A CRC64;

Alignment Scores:

Pred. No.:	5,396-61	Length:	376
Score:	764.50	Matches:	150
Percent Similarity:	84.53%	Conservative:	3
Best Local Similarity:	82.87%	Mismatches:	8
Query Match:	65.68%	Indels:	20
DB:	6	Gaps:	2

US-09-809-545a-1_copy_535_1143 (1-609) x 095K10 (1-376)

```

QY 1 ATGACTAATAAAGCCGCGTAACCCCTACCAATGGCTGGAATTAATCCAGTTGG 60
DB 198 MettrAenLYbThrValaenProTYrThraenGlyTrpLYbLeuAenProValVal 217
QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGCAGCGTCTGTGCGAGGCCAACAG 120
DB 218 GYAlaValaLYSerProGluPheryrAlaGlyThrValaLeuLeuCYsGlnAlaenGln 237
QY 121 GAGGATCTTCATGATGACAGTGCGCCCACTTCACTTGAATATCTTGCATGCTGCG 180
DB 218 GYAlaValaLYSerProGluPheryrAlaGlyThrValaLeuLeuCYsGlnAlaenGln 237
QY 121 GAGGATCTTCATGATGACAGTGCGCCCACTTCACTTGAATATCTTGCATGCTGCG 180
DB 238 GluGlySerSerMetTYrSerAlaProSerSerLeuValTYrThSerAlaMetProGly 257
QY 181 TTTCATATCCGCGCGCCCACTGCTGAGTGCATACCGAGGGGCTCACTTGAGGCCGT 240
DB 258 PheProTYrProAlaAlaThraAlaAlaAlaTYrAGLYAlaAlaLYbLeuArGlyArG 277
QY 241 GGTCCACCGGTGTACAAACCTTACAGAGCTGGCGGCCCAACCCCAATCCCGGCTAT 300
DB 278 GYAlaGThrValTYrAsnThrPheArGAlaAlaAlaProProProProleProAlaTYr 297
QY 301 GCGGAGTAGTGTATACAGCCAGCTGTATGCAATTAATCTTACAGGTTGTTACGCT 360
DB 298 GYgLYAlaValaTYrGlnAspGlyPheTYrGlyAlaAsp---lIeTYrGlyGlyTYrAla 316
QY 361 GCATACCGCTAGCCCGCCAGCCCACTGCTGCTGCTGCTAC----- 405
DB 317 AlaTYrArGTYrAlaGlnProThrProAlaThraAlaAlaTYrSerArpArGlnGln 336
QY 406 -----AGTACAGATTACGAGCAGATT 426
DB 337 PheValaPheValaAlaAlaAspGluLYSerCYsArnThrSer-AlaValaThraAspGluPh 356
QY 427 TATGCTGCCGACCCCTACACACACACTTGTCTCCAGCCCCCACTTACGCGCTGTGCGC 486
DB 356 ewetLeuProThrProThrThrThriSLeuGlnInProProProThraAlaLeuValPr 376
QY 487 A 487
DB 376 o 376

```

RESULT 6

09UGW3

PRELIMINARY:

PRT: 293 AA.

AC 09UGW3. 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE D14IP2.2 (Supported by GENSCAN) (Fragment).
 GN RBM9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049748; CAB63055.1; -.
DR HSSP; P11940; 1CVJ.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 293 AA; 30983 MW; F9A823DD70CC1004 CRC64;

Alignment Scores:
Pred. No.: 5,94e-40 Length: 293
Score: 532.50 Matches: 114
Percent Similarity: 60.65% Conservative: 17
Best Local Similarity: 52.78% Mismatches: 44
Query Match: 45.75% Indels: 41
DB: Gaps: 7

US-09-809-545a-1_copy_535_1143 (1-609) x Q9UGW3 (1-293)
QY 1 ATGACATATAAAGGCGCGTACACCAATGCGTGAATTAATCCAGTTGTG 60
Db 106 MetthrleuylslysmetvalThrProtyrAlaasnlytrpIysleuSerProvalVal 125
QY 61 GCGCGGTCTACACGCCCGCTTATGACGACGCGTGTGTGCGACCAAC- 117
Db 126 GlyAlaValTyrglyProgluLeuTyrrAlaAsnlytrpIysleuSerProvalVal 145
QY 118 ---CAGAGGAGATCTTCATGATACAGTGGCCCGCTTACTT-----GTATATATCTT 168
Db 146 GlyAsnspAlaAlaValProleuSerIyArgIyIleAsnThyTyrIleProleu 165
QY 169 GCAATGCGCTTCTTCATATCCG---GCCGCACCTGCTGACATACCGAGGAGCT 225
Db 166 IleIleProgluPheProtyrProThrAlaAlaThrThrAlaAlaPheArgIyAla 185
QY 226 CACCTTGAGGCGGTGTGCGACCGGTGTAACAACCTTCAGAGCTGCGGCCCGCCACCC 285
Db 186 HistleuArgIyArgIyArgThrValTyrglyAlaValArg---AlaValProProthr 204
QY 286 CCAATCCCGGCTTATGCGGAGTGTATCAAGAGCCAGTGTATGCAATPAATTGCTA 345
Db 205 AlaIleProAlaTyrrProgluIleValIleuInguProIleIleSerAlaIysIlePro 224
QY 346 CAGGTGGTATCGCTGATACCGGTACGCGACCGCCACCCCTGCCACT-----CTT 393
Db 225 GluIyIyIyAlaAlaIyArgIyArgIyArgIyArgIyArgIyArgIyArgIyArgIy 244
QY 394 -----GCTGCTCTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 441
Db 245 AlaAlaAlaAlaAlaAlaAlaIySerAspGlyTyrglyArgValTyrrAlaAspPro 264
QY 442 TACACACACACACTTGTCTCCGCCCCACCTACGCGCTTGTGCTCATAGCTTTTGGC 501
Db 265 Tyr---HisAlaIleuAlaProAlaAlaSerTyrglyValIyAlaAlaVal----- 279
QY 502 CCGTTGACCGATGCAAGACTAGAGACCATGTATGTATGTGGGTCTTCTTCTTCA 561
Db 279 ----- 279
QY 562 TTGACGCTATATATACCAAGGGGATACACCGTTTGTCTCATAT 609
Db 280 -----AlaSerleuTyrrArgIyGlyTyrrSerArgPheAlaProtyr 293
RESULT 7
Q8WYB1 PRELIMINARY; PRT; 380 AA.

AC Q8WYB1;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 2 (RNA binding motif protein 9).
GN HNRBP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229058; AAL71905.1; -.
DR EMBL; BC025281; AAH25281.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PRINTS; PR00308; ANTI-FREEZE1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ SEQUENCE 380 AA; 40351 MW; 601FAE31E7C20715 CRC64;

Alignment Scores:
Pred. No.: 2.43e-39 Length: 380
Score: 526.00 Matches: 116
Percent Similarity: 59.55% Conservative: 15
Best Local Similarity: 52.73% Mismatches: 43
Query Match: 45.19% Indels: 46
DB: Gaps: 8

US-09-809-545a-1_copy_535_1143 (1-609) x Q8WYB1 (1-380)
QY 1 ATGACATATAAAGGCGCGTACACCAATGCGTGAATTAATCCAGTTGTG 60
Db 190 MetthrleuylslysmetvalThrProtyrAlaasnlytrpIysleuSerProvalVal 209
QY 61 GCGCGGTCTACACGCCCGCTTATGACGACGCGTGTGTGCGACCAAC- 117
Db 210 GlyAlaValTyrglyProgluLeuTyrrAlaAlaSerPheGlnAlaAspValSerleu 229
QY 118 ---CAGAGGAGATCTTCATGATACAGTGGCCCGCTTACTT-----CTT 156
Db 230 GlyAsnspAlaAlaValProleuSerIyArgIyIleAsnThyTyrIleProleu 249
QY 157 GTATATATCTTCTGCAATGCTGCTTTCATATCCG---GCCGCACCTGCTGACGTGCA 213
Db 250 IleSerleuProleuValProgluPheProtyrProThrAlaAlaThrThrAlaAla 269
QY 214 TACGAGGGGTCACTTTCAGAGCGCGGTGCGACCGGTGTAACAACACTTCAGAGCTGCG 273
Db 270 PheArgIyAlaHisleuArgIyArgIyArgThrValTyrglyAlaValArg---Ala 288
QY 274 GCGCCCCACCCCAATCCCGGCTATGCGGAGTATGATCAAGCCAGTATAGC 333
Db 289 ValProProThrAlaIleProAlaIyProgluValTyrrAlaAspGlyPheTyrrIy 308
QY 334 AATAAATTGCTACAGGTTGTATGCTGATACCGCTACCGCCACCCCTGCCACT 393
Db 309 AlaAsp---leuTyrglyIyAlaAlaIyArgIyArgIyArgIyArgIyArgIyArgIy 327
QY 394 -----GCTGCTCTTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 429
Db 328 AlaAlaThrAlaAlaAlaAlaAlaAlaAlaIySerAspGlyTyrglyArgValTyrr 347

Qy 430 GCTGCCGACCCCTACACACACTTGTCTCCAGCCCCCAGCTGAGGCTGTGTCATG 489
Db 348 ThrAlaAspProTyr---HisAlaLeuAlaProAlaAlaSerTyrGlyValGlyAlaVal 366
Qy 490 AATGCTTTTGGCCCTTGACCGCATGCGAAGACTAGAGCCATGCTGATGATGTGGTCTC 549
Db 366 ----- 366
Qy 550 GTTCTTTCTTCATTCGAGCTAGTATATACCAAGGGGATACACCGTTTGTCCATAT 609
Db 367 -----AlaSerLeuTyrArgGlyGlyTyrSerArgPheAlaProTyr 380
RESULT 8
Q8TD00 PRELIMINARY; PRT; 390 AA.
ID Q8TD00
AC Q8TD00
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RNA binding motif protein 9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21864496; PubMed=11875103;
RA Norris J.D., Fan D., Sheik A., McDonnell D.P.;
RT "A Negative Coregulator for the Human ER.";
RL Mol. Endocrinol. 16:459-468(2002).
DR EMBL; AY072786; AAL67150.1;
SQ SEQUENCE 390 AA; 41374 MW; 40D6C8C40764B317 CRC64;

Alignment Scores:
Pred. No.: 2,446-39 Length: 390
Score: 526.00 Matches: 116
Percent Similarity: 59.55% Conservative: 15
Best Local Similarity: 52.73% Mismatches: 43
Query Match: 45.19% Indels: 46
Gaps: 8
DB: 4
US-09-809-545A-1_COPY_535_1143 (1-609) x Q8TD00 (1-390)

Qy 1 ATGACTAATAAAAGCCCTGAACCCCTACACCAATGCTGGAAATTAATCCAGTTGTG 60
Db 200 MetTrnAsnLysLysMetValThrProTyrAlaAsnGlyTyrPylLeuSerProValVal 219
Qy 61 GCGCGGCTTACAGCCCGCATCTTATGACAGGACCGTGTGTGTGCCAGGCCAAC--- 117
Db 220 GtAlaValTyrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu 239
Qy 118 ---CAGGAGGATCTTCATGATGACAGTGGCCCACTCA-----CTT 156
Db 240 GtAlaAspAlaAlaValProLeuSerGlyArgGlyGlyIleAsnThrTyrIleProLeu 259
Qy 157 GTATATACTTCGCAATGCTGCTTTCATATCCG---GCCGCCACTGCTGCAGCTGCA 213
Db 260 IleSerLeuProLeuValProGlyPheProTyrProThrAlaAlaThrThrAlaAla 279
Qy 214 TACCGAGGGGCTACCTTGGAGGCGGTGTGCGACCGTGTACACACCTTCAGAGCTGG 273
Db 280 PheArgGlyAlaIleLeuAlaArgGlyArgGlyThrValTyrGlyAlaValArg---Ala 298
Qy 274 GGGCCCCCAGCCCAATCCCGGCTATGGCGAGTAGTATCAAGACAGCAGTATGGC 333
Db 299 ValProProThrAlaIleProAlaTyrProGlyValValTyrIleAsnArgPheTyrGly 318
Qy 334 AATTAATCTACAGGCTGTTACGCTGATACCGCTACGCCAGCCACCCCTGCCACT 393
Db 319 AlaAsp---LeuTyrGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProAlaThrAlaThr 337
Qy 394 -----GCTGCTGCTTACAGTACGATGACGATTACGAGCAGATTAT 429

Db 338 AlaAlaThrAlaAlaAlaAlaAlaAlaAlaTyrSerAspGlyTyrGlyArgValTyr 357
Qy 430 GCTGCCGACCCCTACACACACTTGTCTCCAGCCCCCAGCTGAGGCTGTGTCATG 489
Db 358 ThrAlaAspProTyr---HisAlaLeuAlaProAlaAlaSerTyrGlyValGlyAlaVal 376
Qy 490 AATGCTTTTGGCCCTTGACCGCATGCGAAGACTAGAGCCATGCTGATGATGTGGTCTC 549
Db 376 ----- 376
Qy 550 GTTCTTTCTTCATTCGAGCTAGTATATACCAAGGGGATACACCGTTTGTCCATAT 609
Db 377 -----AlaSerLeuTyrArgGlyGlyTyrSerArgPheAlaProTyr 390
RESULT 9
Q923W8 PRELIMINARY; PRT; 377 AA.
ID Q923W8
AC Q923W8
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative RNA-binding protein fxh.
GN FXH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21294761; PubMed=11401487;
RA Lieberman A.P., Friedlich D.L., Harrison G., Howell B.W., Jordan C.L.,
RA Breddlove S.M., Fischbeck K.H.;
RT "Androgens Regulate the Mammalian Homologues of Invertebrate Sex
RT Determination Genes tra-2 and fox-1.";
RL Biochem. Biophys. Res. Commun. 282:499-506(2001).
DR EMBL; AF387322; AAK4287.1;
DR MGD; MGI:1935973; Fxn.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PS0102; RRM, 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 377 AA; 40165 MW; 249FA5BA6A0D155A CRC64;

Alignment Scores:
Pred. No.: 2,996-39 Length: 377
Score: 525.00 Matches: 116
Percent Similarity: 60.19% Conservative: 14
Best Local Similarity: 53.70% Mismatches: 44
Query Match: 45.10% Indels: 42
DB: 11
US-09-809-545A-1_COPY_535_1143 (1-609) x Q923W8 (1-377)

Qy 1 ATGACTAATAAAAGCCCTGAACCCCTACACCAATGCTGGAAATTAATCCAGTTGTG 60
Db 191 MetTrnAsnLysLysMetValThrProTyrAlaAsnGlyTyrPylLeuSerProValVal 210
Qy 61 GCGCGGCTTACAGCCCGCATCTTATGACAGGACCGTGTGTGTGCCAGGCCAAC--- 117
Db 211 GtAlaValTyrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu 230
Qy 118 ---CAGGAGGATCTTCATGATGACAGTGGCCCACTTCACTT-----GTATATACTTCT 168
Db 231 GtAlaAspAlaAlaValProLeuSerGlyArgGlyGlyIleAsnThrTyrIleProLeu 250
Qy 169 GCATAGCTGCTTTCATATCCG---GCCGCCACTGCTGCAGCTGATACGAGGGGCT 225
Db 251 IleIleProGlyPheProTyrProThrAlaAlaThrThrAlaAlaAlaPheArgGlyAla 270
Qy 226 CACCTTCAGAGCGGTGTGTCAGCGTGTACACACCTTCAGAGCTGCGGCGCCACCC 285
Db 271 HisLeuArgGlyArgGlyArgThrValTyrGlyAlaValArg---AlaValProProThr 289

QY 286 CCAATCCCGCCTATGCGGAGTAGTATCAAGACCATGTATGCGAATTAATTGCTA 345
DB 290 AAlaIleProAlaTyPProGlyValValTyRglInAspGlyPheTyRglAlaAsp---Leu 308
QY 346 CAGGTGTGTTACGCTGATACCGGTACGCCACCCCTGCGACT----- 393
DB 309 TyRglYglYTyRAlaAlaTyRglYTyRglInProAlaThrAlaAlaThrAla 328
QY 394 -----GCTGCTGCTTACAGTATGACGAGGAGTTATGCTGCCAGCC 441
DB 329 AAlaIaAlaAlaAlaAlaAlaTyRserAspGlyTyRglYAlaValTyRThAlaAspPro 348
QY 442 TACCAACACACACTGCTGACGCCGCCACCTACGCGCTTGCTGCATGATGCTTTGGC 501
DB 349 Tyr---HisAlaLeuAlaProAlaAlaSerTyRglYAlaVal----- 363
QY 502 CCTTGACCGATGCCAGACTAGAGCCATGCTATGATGCGGTCTGCTTTCTTCA 561
DB 363 ----- 363
QY 562 TTGCGAGCTAGTATATACAGGGGATATACACCGTTTGTCTCATAT 609
DB 364 -----AlaSerLeuTyRArgGlyTyRserArgPheAlaProTyr 377
RESULT 10
Q8V162 PRELIMINARY; PRT; 377 AA.
ID 08V162
AC 08V162;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 2.
GN FXH OR HRNP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229055; AAL71902.1; -.
DR MGD; MGI:1939373; Fxh.
DR InterPro; IPR00104; Antifreeze 1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PSS0030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 377 AA; 40183 MW; 249E72CAF47C505A CRC64;
Alignment Scores:
Pred. No.: 2.99e-39 Length: 377
Score: 525.00 Matches: 116
Percent Similarity: 60.19% Conservative: 14
Best Local Similarity: 53.70% Mismatches: 44
Query Match: 45.10% Indels: 42
Gaps: 8
DB: 11
US-09-809-545A-1_COPY_535_1143 (1-609) x Q8V162 (1-377)
QY 1 ATGACTATAAAGGCGGTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 191 MetThrIenLysLysMetValThrProTyRAlaAsnGlyTyrPlysLeuSerProValVal 210
QY 61 GCGCGGTCTACAGCCCGCACTTATGACGACGAGTGTGTTGCGACGCAAC--- 117
DB 211 GlyIaValValTyRglYProGlyLeuTyRAlaAlaSerSerPheGlnAlaAspValSerIeu 230
QY 118 ---CAGGAGGATCTTCATGATACGTGCGCCCACTTACTT-----GTAATAATCTTCT 168
DB 231 GlyAsnGlnAlaAlaValProMetSerGlyTyRglYIleAsnThrTyRilapProIeu 250

QY 169 GCATGCTGCTTCCATATCCG---GCCGCACTGCTGCAAGTGCATACCGAGGCGCT 225
DB 251 IleIleProGlyPheProTyRProThrAlaAlaThrAlaAlaAlaPheArgGlyAla 270
QY 226 CACCTTGAGGCGGTGTCGACCGGTATACACCTTCACAGACTGCGGCGCCCAACC 285
DB 271 HisLeuArgGlyArgGlyArgThValTyRglYAlaValArg---AlaValProProThr 289
QY 286 CCAATCCCGCCTATGCGGAGTATGATCAAGACGAGTATGCGAATTAATTGCTA 345
DB 290 AAlaIleProAlaTyPProGlyValValTyRglInAspGlyPheTyRglAlaAsp---Leu 308
QY 346 CAGGTGTGTTACGCTGATACCGGTACGCCACCCCTGCGACT----- 393
DB 309 TyRglYglYTyRAlaAlaTyRglYTyRglInProAlaThrAlaAlaThrAla 328
QY 394 -----GCTGCTGCTTACAGTATGACGAGGAGTTATGCTGCCAGCC 441
DB 329 AAlaIaAlaAlaAlaAlaAlaTyRserAspGlyTyRglYArgValTyRThAlaAspPro 348
QY 442 TACCAACACACACTGCTGACGCCGCCACCTACGCGCTTGCTGCATGATGCTTTGGC 501
DB 349 Tyr---HisAlaLeuAlaProAlaAlaSerTyRglYAlaVal----- 363
QY 502 CCTTGACCGATGCCAGACTAGAGCCATGCTATGATGCGGTCTGCTTTCTTCA 561
DB 363 ----- 363
QY 562 TTGCGAGCTAGTATATACAGGGGATATACACCGTTTGTCTCATAT 609
DB 364 -----AlaSerLeuTyRArgGlyTyRserArgPheAlaProTyr 377
RESULT 11
Q96D26 PRELIMINARY; PRT; 358 AA.
ID Q96D26
AC Q96D26;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Similar to fox-1 homolog (C. elegans).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADRENAL GLAND;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013115; AAL13115.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PSS0102; RRM; 1.
SQ SEQUENCE 358 AA; 37907 MW; BA4B3740C178E441 CRC64;
Alignment Scores:
Pred. No.: 5.54e-39 Length: 358
Score: 522.00 Matches: 115
Percent Similarity: 60.19% Conservative: 15
Best Local Similarity: 53.24% Mismatches: 44
Query Match: 44.85% Indels: 42
Gaps: 8
DB: 4
US-09-809-545A-1_COPY_535_1143 (1-609) x Q96D26 (1-358)
QY 1 ATGACTATAAAGGCGGTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 172 MetThrIenLysLysMetValThrProTyRAlaAsnGlyTyrPlysLeuSerProValVal 191
QY 61 GCGCGGTCTACAGCCCGCACTTATGACGACGAGTGTGTTGCGACGCAAC--- 117
DB 192 GlyIaValValTyRglYProGlyLeuTyRAlaAlaSerSerPheGlnAlaAspValSerIeu 211

```

QY 118 ---CAGAGGAGCTTCCATGTACAGTGGCCCGAGTTCATT-----GTATATACCTTCT 168
      ::::
Db 212 GtAenAmpAlaAlaValProLeuSerGlyArgGlyGlyIleAsnThrTyrlleProLeu 231
QY 169 GCAATGCTGCTGCTTCCATATCCG---GCCGCCACTGCTGCAGCTGCATACCGAGGGGCT 225
      ::::
Db 232 IleIleProGlyPheProTyrrProThrAlaAlaThrThraAlaAlaPheArgGlyAla 251
QY 226 CACCTTCGAGGGCGGTGGTGCACCGGTACACACCTTCAGAGCTGGCGGCCGCCACCC 285
      |||||
Db 252 HisLeuArgGlyArgGlyArgThrValTyrglyAlaValArg---AlaValProProThr 270
QY 286 CCAATCCCGGCTATGCGCGAGTAGTGTATCAAGAGCCAGTGTATGCGCAATAATGCTA 345
      |||||
Db 271 AlaIleProAlaTyrrProGlyValValTyrglnAspGlyPheTyrglyAlaAsp---Leu 289
QY 346 CAGGCTGTTACCTGCATACCGCTACGCCAGCCACCCCTGCGCACTGCT----- 396
      |||||
Db 290 TyrglyGlyTyrrAlaAlaTyrrArgTyrrAlaGlnProAlaThrAlaThrAlaAla 309
QY 397 -----GCTGCTACAGTGCAGTTACGAGAGGAGTTATGCTGCGAGCC 441
      |||||
Db 310 AlaAlaAlaValAlaAlaAlaTyrrSerAspGlyTyrglyArgValTyrrAlaAspPro 329
QY 442 TACCAACACACACTTGTCTCCAGCCGCCCTACCGCGTGTGTCATGAATGCTTTTGGC 501
      |||||
Db 330 Tyr---HisAlaLeuAlaProAlaAlaSerTyrglyValGlyAlaVal----- 344
QY 502 CCCTTGACCGATGCCAAGACTAGAGCCATGCTGATGATGTGGTCTGTTCTTTCTTCA 561
      |||||
Db 344 ----- 344
QY 562 TTGCAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCATAT 609
      |||||
Db 345 -----AlaSerLeuTyrrArgGlyTyrrSerArgPheAlaProTyrr 358

RESULT 12
O96NL7 PRELIMINARY: PRT: 450 AA.
ID O96NL7
AC O96NL7
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CDNA FJ330651 fis, clone D308T200184, highly similar to RNA binding
DE motif protein 9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Niinomiyama K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Futuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshina A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Maehuo Y., Negai K., Isogai T.,
RT "NEO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK05213; BAB70875.1;
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rtm; 1.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 450 AA, 47236 MW, 173030C5SD5CD762 CRC64;

```

```

Best Local Similarity: 52.27% Mismatches: 45
Query Match: 44.59% Indels: 46
DB: 4 Gaps: 8
US-09-809-545A-1_COPY_535_1143 (1-609) x O96NL7 (1-450)

QY 1 ATGACTAATTAAGGCGGCTTAACCCCTACACCAATGCTGGAATTAATTCACCTTGTG 60
      |||||
Db 260 MetThrAsnLysMetValThrProTyrrAlaAsnGlyTrrPylsLeuSerProValVal 279
QY 61 GGGCGGCTCTACAGCCCGGCTTATGACGACGAGTGTGCTGTCGACGACCAAC--- 117
      |||||
Db 280 GlyAlaValTyrglyProGlnLeuTyrrAlaAlaPheSerPheGlnAlaAspValSerLeu 299
QY 118 ---CAGAGGAGCTTCCATGTACAGTGGCCCGAGTTCATT-----CTT 156
      ::::
Db 300 GtAenAmpAlaAlaValProLeuSerGlyArgGlyGlyIleAsnThrTyrlleProLeu 319
QY 157 GTATATACCTTGCATATCGCTTTCATATCCG---GCCGCCACTGCTGCAGCTGCA 213
      ::::
Db 320 IleSerLeuProLeuValProGlyPheProTyrrProThrAlaAlaThrThraAlaAla 339
QY 214 TACGAGGGGCTCACCTTCGAGCGGTGTCGACCGTGTACACACCTTCAGACTGCG 273
      |||||
Db 340 PheArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyrglyAlaValArg---Ala 358
QY 274 GCGCCCGACCCCAATCCCGGCTATGCGAGAGTAGTGTATCAAGAGCCAGTGTATGCG 333
      |||||
Db 359 ValProProThrAlaIleProAlaTyrrProGlyValValTyrglnAspGlyPheTyrgly 378
QY 334 AATAAATGCTACAGGCTGTGCTTACCGCTGATACCGCTACGCCACCCACCTTCCACT 393
      |||||
Db 379 AlaAsp---LeuTyrglyGlyTyrrAlaAlaTyrrArgTyrrAlaGlnProAlaThrAlaThr 397
QY 394 -----GCTGCTGCTACAGTGCAGTTACGAGAGGAGTTAT 429
      |||||
Db 396 AlaAlaThrAlaAlaAlaAlaAlaAlaAlaAlaAlaTyrglyAspGlyTyrglyValTyrr 417
QY 430 GCTGCGACCCCTTACACACACACTTGTCTCCAGCCGCCCTACCGCGTGTGTCATG 489
      |||||
Db 418 ThrAlaAspProTyrr---HisAlaLeuAlaProAlaAlaSerTyrglyValGlyAlaVal 436
QY 490 AATGCTTTTGGCGCCTTGACCGCATGCCAAGACTAGAGCCATGCTGATGTGGTCTC 549
      |||||
Db 436 ----- 436
QY 550 GTTCTTTCTTATGAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCATAT 609
      |||||
Db 437 -----AlaSerLeuTyrrArgGlyTyrrSerArgPheAlaProTyrr 450

RESULT 13
O8VI61 PRELIMINARY: PRT: 303 AA.
ID O8VI61
AC O8VI61
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 3 (Fragment).
OS HNRNP3.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BA1B/C.
RA Chen W., Winkelman J.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229056; AAL71903.1;
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rtm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.

```

Alignment Scores: 1.08e-38 Length: 450
 Score: 519.00 Matches: 115
 Percent Similarity: 58.64% Conservative: 14

DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
FT NON_TER 1 1
SQ SEQUENCE 303 AA; 32864 MW; 74B0AF9F87E935 CRC64;
Alignment Scores:
Pred. No.: 5.27e-33 Length: 303
Score: 456.00 Matches: 95
Percent Similarity: 48.33% Conservative: 6
Best Local Similarity: 45.45% Mismatches: 32
Query Match: 39.18% Indels: 76
DB: 11 Gaps: 4
US-09-809-545a-1_copy_535_1143 (1-609) x Q8V161 (1-303)
QY 7 AATAAAGGCGCGTGAACCCCTACACCAATGCTGAAATTAATTCAGTTGTGGCGCG 66
Db 123 AsnAsnAlaThrAlaAlaGValMetThrAsnGlyTrpIysLeuAsnProValValGlyThr 142
QY 67 GTCTACAGCCCGCAGCTTCTATGACAGGACGGTGTCTGTGCGACGCCAACGAGAGGA 126
Db 143 ValTyrGlyProGluPheTyrAlaValThr----- 152
QY 127 TCTTCATGTACAGTGGCCCGCAGTTCATTAATCTTGCATATGCTGCTTCCA 186
Db 153 -----SerPhePro 155
QY 187 TATCCGCGCCGACGTGTCAGTGCATACCGAGGGGCTGCTGAGCGCGTGTGCGC 246
Db 156 TyrProThrThrGlyThrAlaValAlaTyrArgGlyThrIshLeuArgGlyAlaArg 175
QY 247 ACCGTGTAACAACACCTTCAGAGCTGCGGCGCCGCCAACATCCGCGCTATGCGGA 306
Db 176 AlaValTyrAsnThrPheArgAlaAlaProProProPheProThrTyrGlyAla 195
QY 306 ----- 306
Db 196 AlaLeuGluGlnThrLeuValIshMetProValProTyrAlaGlyLeuAlaProCysPro 215
QY 306 ----- 306
Db 216 LeuProProGlnGlnThrProGluProAlaTyrProThrSerProAlaPheProLeu 235
QY 307 -----GTAAGTATACAGACCGCATGTATGGCAATTAATTGCTA 345
Db 236 SerCysProPheAlaSerArgValValTyrGlnAspGlyPheTyrGlyAlaGlu---1le 254
QY 346 CAGGTGTACCGTGCATACCGCTACGCGCCAGCCACCCCTGCCACTGCTGCTGCTAC 405
Db 255 TyrGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProAlaAlaAlaThrAlaAlaTyr 274
QY 406 AGTACAGTTACGAGACGAGTTAT---GCTGCGACCGCCCTACCAACACAGACTGCTCCA 462
Db 275 SerAspSerTyrGlyArgValTyrAlaAlaAspProTyrIshIshThrIleGlyPro 294
QY 463 GCCCCACCTACGCGCTGTGTCATG 489
Db 295 ThrAlaThrTyrSerIleGlyThrMet 303
RESULT 14
Q8TCMO PRELIMINARY; PRT; 263 AA.
AC Q8TCMO; 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 28.6 kDa protein (Fragment).
GN DKFZP547L059.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1) SEQUENCE FROM N.A.

RC TISSUE=BRAIN;
RA Bloeker H., Boecker M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL13700; CAD28499.1; -
KW Hypothetical protein.
FT NON_TER 263 263
SQ SEQUENCE 263 AA; 28565 MW; 56C2D853846BCE1 CRC64;
Alignment Scores:
Pred. No.: 1.75e-23 Length: 263
Score: 351.00 Matches: 65
Percent Similarity: 97.06% Conservative: 1
Best Local Similarity: 95.59% Mismatches: 2
Query Match: 30.15% Indels: 0
DB: 4 Gaps: 0
US-09-809-545a-1_copy_535_1143 (1-609) x Q8R2T5 (1-435)
QY 1 ATGACTAATAAAGGCGGTGAACCCCTACACCAATGCTGAAATTAATTCAGTTGTG 60
Db 196 MetThrAsnIshTyrThrValAsnProTyrThrAsnGlyTrpIysLeuAsnProValVal 215
QY 61 GCGCGGTCTACAGCCCGCAGCTTCTATGACAGGACGGTGTCTGTGCGACGCCAACAG 120
Db 216 GlyAlaValTyrSerProGluPheTyrAlaGlyThrValLeuIshCysGlnAlaAsnGln 235
QY 121 GAGGATCTTCATGTACAGTGGCCCGCAGTTCATTAATCTTGCATATGCTGCGC 180
Db 236 GluIshSerIshTyrSerAlaPheSerSerLeuValTyrThrSerAlaMetProGly 255
QY 181 TTTCATATCCGCGCGCAGTGTCT 204
Db 256 PheProTyrProAlaAlaThrAla 263
RESULT 15
Q8R2T5 PRELIMINARY; PRT; 435 AA.
AC Q8R2T5; 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Similar to RNA binding motif protein 9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1) SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027263; AAH27263.1; -
SQ SEQUENCE 435 AA; 46252 MW; 62A355605FF8A5B4 CRC64;
Alignment Scores:
Pred. No.: 1.01e-18 Length: 435
Score: 299.00 Matches: 80
Percent Similarity: 53.94% Conservative: 9
Best Local Similarity: 48.48% Mismatches: 55
Query Match: 25.69% Indels: 21
DB: 11 Gaps: 7
US-09-809-545a-1_copy_535_1143 (1-609) x Q8R2T5 (1-435)
QY 1 ATGACTAATAAAGGCGGTGAACCCCTACACCAATGCTGAAATTAATTCAGTTGTG 60
Db 259 MetThrAsnIshTyrThrValAsnProTyrThrAsnGlyTrpIysLeuAsnProValVal 278
QY 61 GCGCGGTCTACAGCCCGCAGCTTCTATGACAGGACGGTGTCTGTGCGACGCCAAC---117
Db 279 GlyAlaValTyrGlyProGluLeuTyrAlaAlaSerIshGlnAlaAspValSerLeu 298
QY 118 ---CAGAGGAGATCTTCATGTACAGTGGCCCGCAGTTCATTTCT 168

```

Db 299 GlyAenGIuAlaAlaValProleuSerGIyArgGIyGlyIleAsnThrTyrlleProleu 318
Qy 169 GCAATGCTGCTGCTTTCATATCCG--GCCGGCACTGCTGCAGCTGCATACCGAGGGGCT 225
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 IleIleProGIyPheProTyPProThrAlaIleThrAlaIleAlaIlePheArgGIyAla 338
Qy 226 CACCTTCAGGCGCGTGTGCGCACCGGTACAAACACCTTCAGAGCTGCCGCGCCACCC 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 HisLeuArgGIyArgGIyArgThrValTyrgIyAlaValArg--AlaValProProThr 357
Qy 286 CCAATCCCGGCTATGGGGAGTAGTATCAAGAGCCAGTGTATGGCAATAAATTGCTA 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 AlaIleProAlaTyPProGIyValAspMetGln--ProThrAspMetHisSerIleuLeu 376
Qy 346 CAGGGGTGTTACGCTGCATACCGCTACGCCAGCCACCCCTGCCACTGCTGCTGCT-- 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 -----LeuGlnProGlnPro-GlnLeuLeuGlnProle 387
Qy 404 -----ACAGTGACAGTTACGAGACGAGTTTATGCTGCCGACCCCTACCGACACACA 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 uGlnProleuThrAlaThrValThrAlaGIyCySerThrGlnLeuThrProThrMetProSe 407
Qy 454 CTGCTCCAGCCC 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 rProleuProPro 411

```

Search completed: March 12, 2003, 22:23:32
 Job time : 43.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 12, 2003, 21:18:34 ; Search time 13.5 Seconds
(without alignments)
2654.601 Million cell updates/sec

```
Title: US-09-809-545A-1_COPY_535_1143
Perfect score: 1164
Sequence: 1 atgactaatataaaagcgct.....acaacggtttgtctcatat 609
```

Scoring table:		BLOSUM62
Xgapop	10.0	Xgapext 0.5
Ygapop	10.0	Ygapext 0.5
Fgapop	6.0	Fgapext 7.0
Delop	6.0	Delext 7.0

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 525148

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```

Command line parameters:
MODEL=frame n2p.model -DEV=x1h
O=/gen2.1/USPTO.spool/US09080545/rnat 110322003 110947 12341/app.qcut.fasta-1.775
DB=issued patents A4 -OFTM=fastan -SUFFIX=12p.ra1 -MIMATCH=0.1 -LOOPEC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsomed2 -TRANS=align40.cdd
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINTEN=0 -MAXLEN=2000000000
-USER=US9809545 @CCN 1 1.13 @rnat 11032003 110947 12341 -NCPU=6 -ICPU=3
NO_XLPTX -NO_MAP -LARGEBUFFER -NEG_SCORES=0 -WAIT -LONGLOG -DE/ TIMEOUT=120
WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

```
Database : Issued Patients AA:*
1 : /cgn2.6/prodata/1/1aa/5A_COMB.pep.*
2 : /cgn2.6/prodata/1/1aa/5F_COMB.pep.*
3 : /cgn2.6/prodata/1/1aa/6A_COMB.pep.*
4 : /cgn2.6/prodata/1/1aa/6E_COMB.pep.*
5 : /cgn2.6/prodata/1/1aa/PTCUTS_COMB.pep.*
6 : /cgn2.6/prodata/1/1aa/backfile1.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	624	53.6	330	4	US-09-145-391-2	Sequence 2, Appli
2	315.5	27.1	366	3	US-09-176-657-2	Sequence 2, Appli
3	100.5	8.6	349	4	US-09-343-018-1	Sequence 1, Appli
4	96	8.2	760	1	US-08-195-152-2	Sequence 2, Appli
5	95.5	8.2	1172	1	US-08-313-288B-19	Sequence 19, Appli
6	94.5	8.1	1964	4	US-03-461-997-1	Sequence 1, Appli
7	93	8.0	444	4	US-09-252-292C-27	Sequence 27, Appli
8	92.5	7.9	362	1	US-08-437-027-21	Sequence 21, Appli
9	92.5	7.9	365	1	US-08-437-027-20	Sequence 20, Appli
10	92.5	7.9	656	2	US-08-343-443B-2	Sequence 2, Appli
11	92.5	7.9	656	4	US-03-214-564A-4	Sequence 4, Appli
12	92	7.9	219	4	US-09-521-345-2	Sequence 2, Appli

13	91	7.8	3.6	4	US-09-343-011B-2	Sequence 2, Appl.
14	87.5	7.5	980	4	US-09-443-100-8	Sequence 8, Appl.
15	87	7.5	379	1	US-08-455-142A-11	Sequence 11, Appl.
16	87	7.5	1706	2	US-08-459-568-2	Sequence 2, Appl.
17	87	7.5	1706	2	US-08-399-411-2	Sequence 2, Appl.
18	87	7.5	1706	3	US-08-516-859A-2	Sequence 2, Appl.
19	87	7.5	1706	4	US-09-586-472-2	Sequence 2, Appl.
20	87	7.5	1706	4	US-09-528-706-2	Sequence 2, Appl.
21	86.5	7.4	262	3	US-08-946-914-14	Sequence 14, Appl.
22	86.5	7.4	262	4	US-09-656-450-14	Sequence 14, Appl.
23	86.5	7.4	1274	4	US-09-059-443-2	Sequence 2, Appl.
24	86.5	7.4	4302	3	US-08-658-136-5	Sequence 8, Appl.
25	86.5	7.4	4302	4	US-09-052-463-8	Sequence 8, Appl.
26	86.5	7.4	4339	4	US-09-052-463-6	Sequence 8, Appl.
27	86	7.4	207	4	US-09-336-536-50	Sequence 50, Appl.
28	86	7.4	245	4	US-09-336-536-48	Sequence 48, Appl.
29	86	7.4	462	3	US-08-946-944B-5	Sequence 5, Appl.
30	86	7.4	462	4	US-09-116-049-4	Sequence 4, Appl.
31	85.5	7.3	469	1	US-08-313-288B-15	Sequence 15, Appl.
32	85.5	7.3	572	6	5256770-7	Patent No. 5256770
33	85.5	7.3	1185	5	US-09-041-886-23	Sequence 23, Appl.
34	85	7.3	275	1	US-08-4312-870-7	Sequence 7, Appl.
35	85	7.3	275	1	US-08-431-387-3	Sequence 3, Appl.
36	85	7.3	275	1	US-08-322-677A-7	Sequence 7, Appl.
37	85	7.3	275	1	US-08-322-676-7	Sequence 7, Appl.
38	85	7.3	275	1	US-08-460-343B-74	Sequence 74, Appl.
39	85	7.3	275	1	US-08-398-028B-74	Sequence 74, Appl.
40	85	7.3	275	2	US-08-504-265B-90	Sequence 90, Appl.
41	85	7.3	275	2	US-08-140-083A-9	Sequence 9, Appl.
42	85	7.3	275	2	US-08-865-203-8	Sequence 8, Appl.
43	85	7.3	275	2	US-09-135-658-3	Sequence 3, Appl.
44	85	7.3	275	2	US-07-849-420-8	Sequence 8, Appl.
45	85	7.3	275	3	US-08-898-218-7	Sequence 7, Appl.

ALIGNMENTS

```

RESULT 1
US-09-145-391-2
; Sequence 2, Application US/09145391
; Patent No. 6194171
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/145,391
; CURRENT FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-145-391-2

```

Alignment Scores:	
Pred. No.:	1.6e-54
Score:	624.00
Percent Similarity:	88.78%
Best Local Similarity:	86.86%
Query Match:	5.61%
Da:	4
Length:	330
Matches:	119
Conservative:	4
Mismatches:	12
Gaps:	2
Indels:	0

US-09-809-545A-1_COPY_535_1143 (1-609) X US-09-145-391-2 (1-330)

```
QY      1 ATGACTATATAAAGACCGCGTGAACCCCAKCAACAATGGCGGAATTAAATCAAGTTGTG   60
Db      196 MetThrAsnLysLysThrValAsnProCylrThraSngLyIrrPlysLeaAsnPrivaVal    215
QY      61 GGGGGGCGCTACAGCCCCGACCTTCATATGCAGAGCACGGTGCTGTGTGCCAGAGCCAACAG   120
        |||||:::|||||
```

Db	216	GLYALVALTYRSEPRGGLPHETHYALAEIYLHYRVALLEUCYSGINALAENGIN	235
Qy	121	GAGGAGATTCCTCAGTACAGGCGCCAGTCACTGTATATCTTGCAATGCCGCG	180
Db	236	GIUGYSESRSEMTYRSEALAPROSESRSEUVALTYRMRSEALAKEPROGLY	255
Qy	181	TTTCCATATCCGGCGCCACTGCTGCACGTGCATACCGAGGGGCTACCTTGAGCGGT	240
Db	256	PHETROYRPROALAAETHRALAIALAIALATYRGGIYALAHISLEUAHGGIYARG	275
Qy	241	GGTGGCACCGGTACAACACTTCAGAGCTCGGGCGGCCACCCCAATCCGGGCTAT	300
Db	276	GIYAGHTRVALTYRAGHTRHHPHEAGHIALAIALAPROFROPOLLEPROALATYR	295
Qy	301	GGCGGAGTAGTGATCAAGAGCCAGGTGATGCAATAAATTGCTACAGGGGTGTACGCT	350
Db	296	GLIYGLVALIATYRPROGLIYTRPILREUHTPRCYBARHISLEU--TRTPRLEUCYSC	315
Qy	361	GCATACGCGTACGGCGGACCCAGCCCTGCGCACTGCTGCTGCTACAG	407
Db	315	YELLEROLEUARGPROALATYRTPCYAHISVCYBARCYBLEGIN	330

```

1 RESULT 2
2 US-09-176-657-2
3
4 : Sequence 2, Application US/09176657
5 : Patent No. 6020164
6 :
7 : GENERAL INFORMATION:
8 :   APPLICANT: Bandman, Olga
9 :   APPLICANT: Tang, Y. Tom
10 :  APPLICANT: Corley, Neil C.
11 :  APPLICANT: Guegler, Karl J.
12 :
13 :  APPLICANT: Lu, Alina
14 :
15 :  APPLICANT: Baughn, Mariah R.
16 :
17 :  TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
18 :
19 :  FILE REFERENCE: PF-0611 US
20 :
21 :  CURRENT APPLICATION NUMBER: US/09/176,657
22 :
23 :  CURRENT FILING DATE: 1998-10-21
24 :
25 :  NUMBER OF SEQ ID NOS: 9
26 :
27 :  SOFTWARE: PERL Program
28 :
29 :  SEQ ID NO 2
30 :
31 :    LENGTH: 366
32 :    TYPE: prt
33 :
34 :  ORGANISM: Homo sapiens
35 :
36 :  FEATURE: -
37 :
38 :  OTHER INFORMATION: 1250374
39 :
40 :  US-09-176-657-2

```

Alignment Scores:	
Pred. No.:	1,41e-23
Score:	315.50
Percent Similarity:	49.05%
Best Local Similarity:	42.86%
Query Match:	27.10%
DB:	3
Length:	366
Matches:	90
Conservative:	13
Mismatches:	59
Indels:	10
Gaps:	48

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-176-657-2 (1-366)

QY 1 ATCTGATATAAAGAGCGCGTGAACCCCTGACACCAATGGCTGGAAATTAATTCAGTTGTG 60
Db MethrinenblybMetValThrProtyrAlaasnGlyTrpIlybLseuSeProValVal 209
QY 61 GGC CGCGCTCTACACCCCGCATCTTCATGCACAGACGGGTGCTGTGTCACAGGCCAAC-- 11.7
Db 210 GlyAlaValIlyrGlyProGluLeuIlyrAlaIalaserSerPheGlnAlaApValSerLeu 229
QY 118 --GAGGAGGATCTTCCATGTACAGTGCAGGCCCGCCAGTTCACT-----GTATATCTCT 168
Db 230 GlyAsnAepAlaAlaValProLseuSerGlyrGlyGlyLeuThrIlyrLeuProLseu 249
QY 169 GCATGCTGCTGGCTTCCATATCG---CGCGCACTGCTGTCAGCTGATACCGAGGGCT 225
Db 250 IleIleProGlyPheProIyProthrAlaIalThrAlaAlaAlaPheAsrGlyAla 269

Qy	226	CACCTTGAGAGCCGTGTCTCCACCGTTACAACACCTTCAGACTGGCGGCCCCACCC	2855
Db	270	HisLeuArgGlyArgGlyAlaGlnThrValTyrGlyAlaValArg---	AlaValProProThr 2888
Qy	286	CCAATCCCGGCTATGGCGGAGTAGTGATCAAGAGCCAGTGTATGGCAATTAATTGCTA	3455
Db	289	AlaIleProAlaTyrProGlyValAspMetGln---	ProThrAspMetHisSerLeu 3070
Qy	346	CAGGGTGGTAACTGTGCATACCGCTACGCCACCCACCCCTGCCAATGTGCTGCTCT--	4030
Db	308	-----LeuGlnProIlePro--ProLeuLeuGlnProLe	318
Qy	404	-----ACAGTGCAGTTACGAGCAGAGATTATGCTCCGACCCCTCAACACACACA	4530
Db	318	uGlnProLeuThrValThrValMetAlaGlyCysThrGlnProThrProThr-----	3355
Qy	454	CTTGCTCCAGCCCCCACACTACGGCGTTGTGCATGAATGCTTTGGCGCCCTTGACCGAT	5130
Db	336	-----MetProLeuProLe	3400
Qy	514	GCCAAAGCTAGAGACCACATGCTGATGATGTGGGTCTGTTCTTTCTTCATTCAGAGCTAGT	5730
Db	340	uPro---LeuAlaMetGluLeuAlaLeuTyr-----	ArgVal 3510
Qy	574	ATATACCAAGGCGGATCAACACGTTTG	601
Db	351	IlyrThrGluValAlaThrIleAspLeu	360

```

RESULT 3
US-09-343-011B-1
Sequence 1, Application US/09343011B
Patent No. 6300473
GENERAL INFORMATION:
APPLICANT: Stephanie Richard
TITLE OF INVENTION: SLM-1 AND SLM-2; NOVEL
FILE REFERENCE: A32561
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: CA 2265271
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO. 1
LENGTH: 349
TYPE: PRT
ORGANISM: Mus musculus
US-09-343-011B-1

```

Pred. No.:	0.0501	length:	349
Score:	100.50	Matches:	56
Percent Similarity:	35.92%	Conservative:	18
Best Local Similarity:	27.18%	Mismatches:	72
Query Match:	8.63%	Indels:	6
DB:	4	Gaps:	9

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-343-011B-1 (1-349)

Oy	AAATCCAGTTGTGGGCGCGCTACAGCCCCGACTTGA-----	86
Db	GluserglyArgglyIleargglyAArgglyIlethrProthraPro	2020
Oy	-----TCAGGCACGGTGCTGTTGTGCCAGGCCAACGAGGAGATTCCATGTACAG	140
Db	SerArgglyArgglyAlaValProProProPro-Pro-GlyArgglyValLeuth	2220
Oy	TGGCCCCAGTTCACTGTGTATAACTTTCGAAATGCT-----GGCTT	182
Db	rProArgglyThrThraValThrArgglyAlaLeuProvalProProIlealaArgglyVa	2420
Oy	TCCATATCCGCGCCCACTGCTGCAGCTGCATACCAGGGGCTCACCCTTCGAGGCCGTGG	2420


```

Db 242 lProthPrroarglAaarglyThrAlaValProgly----- 255
QY 243 TCGCACCCTGACAACTTCAGAGCTGCGCGCCCAACCCCAATCCCGCTTAGG 302
Db 256 -----TyrArg-----AlaProProProProAlaHisAspAlaTyrGI 268
QY 303 CGGAGTAGTATACAGAGCCAGTGTATGCGAATAAATTGCTACAGGGGTGCTGC 362
Db 268 uGIuTyrGIuTyrAspAspGIuTyrGIuTyrGIuTyrAspAspGIuTyrGIuAl 287
QY 363 ATACCGCTAGCCGACCCCACTGCGCAGCTGCGCTTACAGTACATTAAGGACG 422
Db 287 aTyraAspAspSerTyrValThrProThrGlnSerValProGIuTyrTyrAspTyrGIuHi 307
QY 423 AGTTATGCTGCGCAGCCCTTACACACACACTGTCTCAGCCCACTAGCGGCTGG 482
Db 307 sGIyValAsnGlnAspAlaTyr----- 314
QY 483 TGCCATGATGCTTTTGGCCCTTACCGATGCCAAGTACAGGATGATGATGT 542
Db 315 -----AspSerTyrAlaProGIuTyrAlaThrThrArg----- 326
QY 543 GGGCTGCTTTCTTCTTCATGTCAGGCT-----AGTATATACCAAGGGGATACAA 593
Db 327 -----SerSerLeuLysAlaProProProArgSerAlaArgGIuTyr-- 341
QY 594 CCGTTTGTCTCATAT 609
Db 342 -ArgGlnHisProTyr 346

RESULT 4
US-08-195-152-2
; Sequence 2, Application US/08195152
; Patent No. 5679541
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,152
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-195-152-2

```

```

Alignment Scores:
Pred. No.: 0.182 Length: 760
Score: 96.00 Matches: 42
Percent Similarity: 37.29% Conservative: 24
Best Local Similarity: 23.73% Mismatches: 61
Query Match: 8.25% Indels: 50
DB: 1 Gaps: 6

US-09-809-545a-1_copy_535_1143 (1-609) x US-08-195-152-2 (1-760)
QY 124 GGATCTTCATATGAC-----ACTGGCCCAAGTCACTT 156
Db 223 GIISeasennuTyrGIuYCySSerSerAlaSerAnProLeuAspGIyAlaValAla 242
QY 157 GTATATCTTCTGCAATGCTGCTTTCATATCCGCGCCGCTGCTGACGTGATAC 216
Db 243 ValAsnSerSerAlaVal-----AlaAlaAlaAlaAlaValTyr 256
QY 217 CGAGGGCTCACTTCGAGCGCTGTGCGACCGGTACAAACACTTCAGAGCTGGCGG 276
Db 257 AspGIuYHis-----AspTyrTyrTyrTyrAsnSerMetGlnGlnTyrThr 272
QY 277 CCCCCACCCCAATCCCGCTATGCGGAGTAGTATCAAGCCAGCTATGCAAT 336
Db 273 ProProProPheTyrSerGIuTyrGIuTyrProTyrAlaAlaAlaThrAlaAlaArgGln 292
QY 337 AAATTGCTACAGGGGTGTTAGCTGATACCGCTACCGCCGACCCCTGACACTGCT 396
Db 293 AlAlayMetGIuProGIuAlaAlaAlaAlaAlaAlaTyrLeuThrProSerTyrAla 312
QY 397 GCTGCC----- 402
Db 313 AlaSerGIuAsnAsnSerGlnLeuTyrSerSerProTyrAlaGIuTyrAsnAsnPhe 332
QY 403 -----TACAGTACAGTATACGACAGATTAATGCTGCCGAC 438
Db 333 GLYGlnGlnAspTyrGIuTyrTyrAsnGlnGlnTyrGIuAsnTyrTyrSerProAla 352
QY 439 CCTACACACACACACTGCTCCAGCCCT-----ACCTAGCGGCTTGT--- 483
Db 353 AsnTyrSerProTyrAlaValAlaSerSerProSerSerSerAlaSerHisGIyHisGIyPhe 372
QY 484 GCCATGATGCTTTTGGCCCTTACCGATGCCAAGCTACAGCCATGCT 534
Db 373 HisValAlaAlaSerSerAsnLeuSerGIuSerProThrAspThrHisSer 389

RESULT 5
US-08-313-288B-19
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.

```


QY 322 CCACTGATGCAATTAATTGCTACAGGCTGCTACCGCTACCGCCAGCC 381
DB 89 ProvalGlnGlyTyr-----GlyThrGlyAlaTyrAspThrThrAla 103
QY 382 ACCCTGCACTGCTGCTGCTACAGTACAGTACGACAGAGTTTACGCGACCC 441
DB 104 ThrValThrThrThrGlnAla-----SerTyrAlaGlnSer 116
QY 442 TACACACCACTGCTGCTCCAGCCCACTACGCGCTGGTGCATGATGCTTTGCG 501
DB 117 AlaTyrGlyThrGlnProAlaTyrProAlaTyrGlyGln-----Gln 130
QY 502 CCCTTGACCGATGCCAGACAGTACGACCATGATGATGCTGCTGCTTTCTTCA 561
DB 131 ProAlaAlaThrAlaProThrArgProGlnAspGlyAsnLysProThrGluThrSerGln 150
QY 562 TTGCAGGCTAGTATATACCAAGGGGATACCAACCT 597
DB 151 ProGlnSerSer-----ThrGlyGlyTyrAsnGln 160

RESULT 9

US-08-437-027-20
Sequence 20, Application US/08437027
Patent No. 5670317
GENERAL INFORMATION:
APPLICANT: Landanyi, Marc
APPLICANT: Gerald, William
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/437,027
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 46416/JPM/CCA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-437-027-20

Alignment Scores:

Pred. No.: 0.323 Length: 365
Score: 92.50 Matches: 18
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 7.95% Indels: 41
DB: 1 Gaps: 9

US-09-809-545a-1_copy_535_1143 (1-609) x US-08-437-027-20 (1-365)

QY 85 TATGACGACGACGCTGCTTTGTCAGGCC---AACACAGAGGATGTTTCATGTACAGT 141

DB 29 TyrAlaGlnThrThr-----GlnAlaTyrGlyGlnGlnSerTyrGlyThrTyrGly 45
QY 142 GGGCCGATTCATCTGTAATATCTTGTACAGTGCCTGCTTTCATATCCGGCCCACT 201
DB 46 GlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThrTyrGlyGlnThrAla 65
QY 202 GCTGACGCTGATACCGAGGGGCTCACCCTTGAGAGCCGTGTCGACCCGTATACAAACC 261
DB 66 TyrAlaThrSerTyr-----GlyGlnProProThrGlyTyrThrThr 79
QY 262 TTCAGAGCTGGGGGCGCCCAATCCGCGCTATGCGGAGTATGATATCAAGAG 321
DB 80 -----ProThrAlaProGln-----AlaTyrSerGln 88
QY 322 CCACTGATGCAATTAATTGCTACAGGCTGCTACGCTGATACCGTACCGCCAGCC 381
DB 89 ProvalGlnGlyTyr-----GlyThrGlyAlaTyrAspThrThrAla 103
QY 382 ACCCTGCACTGCTGCTGCTACAGTACAGTACGACAGTATGCTGCGACCC 441
DB 104 ThrValThrThrThrGlnAla-----SerTyrAlaAlaGlnSer 116
QY 442 TACACACCACTGCTGCTCCAGCCCACTACGCGCTGGTGCATGATGCTTTGCG 501
DB 117 AlaTyrGlyThrGlnProAlaTyrProAlaTyrGlyGln-----Gln 130
QY 502 CCCTTGACCGATGCCAGACAGTACGACCATGATGATGCTGCTGCTTTCTTCA 561
DB 131 ProAlaAlaThrAlaProThrArgProGlnAspGlyAsnLysProThrGluThrSerGln 150
QY 562 TTGCAGGCTAGTATATACCAAGGGGATACCAACCT 597
DB 151 ProGlnSerSer-----ThrGlyGlyTyrAsnGln 160

RESULT 10

US-08-343-443B-2
Sequence 2, Application US/08343443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Weloc, Thomas
APPLICANT: Peter, Martine
APPLICANT: PloougasTel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Weiser, Gerard J.
;   REGISTRATION NUMBER: 19,763
;   TELEPHONE: 215-875-8383
;   TELECOMMUNICATION INFORMATION: 989, 6121P
;   TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 656 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-343-443B-2

Alignment Scores:
Pred. No.: 0.39 Length: 656
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 7.95% Indels: 41
DB: Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x US-08-343-443B-2 (1-656)
QY 85 TATGACGAGCAGCGCTGTGTGTCAGGCC---AACCAGAGGAGATCTTCATATACAGT 141
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 TATGAGlnThrThr-----GlnAlaTyrglynglnseryTyrglyThTyrgly 45
QY 142 GGGCCGACGTCTGTATATACCTTGTGCAATGCGCTTCCATATCCGGCCGCACT 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 46 GlnProThrsPvalserTyThrglnAlaGlnThrThralaThTyrglyGlnThrala 65
QY 202 GCTGACCTCATACCGAGGGGCTCACCTTCGAGCGCTGCTGCACCGTGTACACACC 261
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 66 TyralaThrseryTy-----GlyGlnProThrglyTyThrThr 79
QY 262 TTCAGAGCTGCGGCGCCCAACCCCAATCCGGCTATGGCGAGTATCAAGAG 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 80 -----ProThralaProGln-----AlaTySergln 88
QY 322 CCAGTATGCAATAATGTCTACAGGCTGTACGCTGATACCGCTACGCCAGCC 381
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 ProvalGlnTyTy-----GlyThrglyAlaTyraPThrThrThrala 103
QY 382 ACCCTGCACTGCTGCTGTCTACAGTACAGATTACGAGAGTTTATGCTGCCAGCCC 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 104 ThValThrThrThrglnAla-----SeryTyraAlaGlnSer 116
QY 442 TACCACCAACACTGTCTCCAGCCCCCACTACGCGCTGTGTCATGATGCTTTGCG 501
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 AlaTyrglyThrglnProAlaTyProAlaTyrglyGln-----Gln 130
QY 502 CCCTTGACCATGCGACAGTACGAGCATGATGATGAGGTCTGCTCTTTCTTCA 561
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131 ProAlaAlaThralaProThraPProGlnAspGlyasnTyProThrglnThSergln 150
QY 562 TTGCAGGCTAGTATATCAAGGGGAGTACACCGT 597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 151 ProGlnSerser-----ThrglyGlyTyraSngln 160

RESULT 11
US-09-214-564A-4
; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.
; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
; FILE REFERENCE: Elongation By HIV-1 TAT
```

```
; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-4

Alignment Scores:
Pred. No.: 0.39 Length: 656
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 7.95% Indels: 41
DB: Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-214-564A-4 (1-656)
QY 85 TATGACGAGCAGCGCTGTGTGTCAGGCC---AACCAGAGGAGATCTTCATATACAGT 141
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 TATGAGlnThrThr-----GlnAlaTyrglynglnseryTyrglyThTyrgly 45
QY 142 GGGCCGACGTCTGTATATACCTTGTGCAATGCGCTTCCATATCCGGCCGCACT 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 46 GlnProThrsPvalserTyThrglnAlaGlnThrThralaThTyrglyGlnThrala 65
QY 202 GCTGACCTCATACCGAGGGGCTCACCTTCGAGCGCTGCTGCACCGTGTACACACC 261
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 66 TyralaThrseryTy-----GlyGlnProThrglyTyThrThr 79
QY 262 TTCAGAGCTGCGGCGCCCAACCCCAATCCGGCTATGGCGAGTATCAAGAG 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 80 -----ProThralaProGln-----AlaTySergln 88
QY 322 CCAGTATGCAATAATGTCTACAGGCTGTACGCTGATACCGCTACGCCAGCC 381
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 ProvalGlnTyTy-----GlyThrglyAlaTyraPThrThrThrala 103
QY 382 ACCCTGCACTGCTGCTGTCTACAGTACAGATTACGAGAGTTTATGCTGCCAGCCC 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 104 ThValThrThrThrglnAla-----SeryTyraAlaGlnSer 116
QY 442 TACCACCAACACTGTCTCCAGCCCCCACTACGCGCTGTGTCATGATGCTTTGCG 501
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 AlaTyrglyThrglnProAlaTyProAlaTyrglyGln-----Gln 130
QY 502 CCCTTGACCATGCGACAGTACGAGCATGATGATGAGGTCTGCTCTTTCTTCA 561
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131 ProAlaAlaThralaProThraPProGlnAspGlyasnTyProThrglnThSergln 150
QY 562 TTGCAGGCTAGTATATCAAGGGGAGTACACCGT 597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 151 ProGlnSerser-----ThrglyGlyTyraSngln 160

RESULT 12
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Shepherd, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZS163 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
```

;; CURRENT FILING DATE: 1999-03-17
;; PRIOR APPLICATION NUMBER: US 60/124,820
;; PRIOR FILING DATE: 1999-03-17
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 219
;; TYPE: PR1
;; ORGANISM: Homo sapiens
US-09-527-345-2

Alignment Scores:
Pred. No.: 0.307 Length: 219
Score: 92.00 Matches: 53
Percent Similarity: 33.91% Conservative: 6
Best Local Similarity: 30.46% Mismatches: 56
Query Match: 7.90% Indels: 59
DB: Gaps: 10

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-527-345-2 (1-219)

QY 24 CCCCTACACCAATGGCTGGAATTAATCAAGTTGGCGGCTGCTACAGCCGCACTT 83
DB 33 ProlenHisProSerLeuAsn1LeProTyrGlyLe---ArgAsnLeuProProleu 51
QY 84 CTATGACGACGAGTGTCTGCTGCGCCAGCCAAACGAGAGGAGATCTTCATGATACAGTGG 143
DB 52 TyrTyrArgProValAsnThrValProSerTyrProGlyAsnThrTyrThrAspThrG1 71
QY 144 CCCGATTCACCTGATATACTTCTGCAATGCTGCTTCATATCGGCGCCGCACTGC 203
DB 71 YLeuProSerTyrProTyrP1LeuThrSerProGlyPheProTyr--- 86
QY 204 TGCAGCTGCATACCGAGGGGCTCAGCTTCGAGGC----- 237
DB 87 -----ValTyr-----His1LeArgGlyPheProLeuAlaThrGlnLeuAsnVa 101
QY 238 -----CGTGTGCGCACCGCTGAC-----AACAGCTTCGAGCTGC 272
DB 101 LProProleuProProAlaGlyPheProPheValProProSerArgPhePheSerAlaAl 121
QY 273 GCGCGCCCA-----CCCCCAATCCCGCCCTATAGCGGAGTAGTGTATCAAGAGCCACT 326
DB 121 AAla1aProAla1a1aProPro1a1a1a1a-----GluProAl 134
QY 327 GATGCGCAATTAATGCTACAGAGGGGTGTAAGCTGATACCGCTACGCGCAGCCACC 386
DB 134 a-----AlaAla1aProleuth 140
QY 387 TGCCACTGCTGCTGCTACAGTGAAGTACGAGAGATTATGCTGGAGCCCTACCA 446
DB 140 rAla1aThrProValAla1a1aGluProAla1a1aGlyAlaProValAla1a1aGluPro----- 158
QY 447 CCACACACTGCTCCAGCCGCCCACTACCGCGCTGTGCTCC 486
DB 159 -----AlaAlaGluAlaPro-----ValGlyAla 166

RESULT 13
US-09-343-011B-2
; Sequence 2, Application US/09343011B
; Patent No. 6300473
; GENERAL INFORMATION:
; APPLICANT: Stephane Richard
; TITLE OF INVENTION: SLM-1 AND SLM-2: NOVEL
; TITLE OF INVENTION: SAME-1 LIKE MAMMALIAN PROTEINS
; FILE REFERENCE: A32561
; CURRENT APPLICATION NUMBER: US/09/343,011B
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: CA 2265271
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2

;; LENGTH: 346
;; TYPE: PR1
;; ORGANISM: Mus musculus
US-09-343-011B-2

Alignment Scores:
Pred. No.: 0.448 Length: 346
Score: 91.00 Matches: 32
Percent Similarity: 41.67% Conservative: 8
Best Local Similarity: 33.33% Mismatches: 46
Query Match: 7.82% Indels: 10
DB: Gaps: 4

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-343-011B-2 (1-346)

QY 178 GGCCTTCATATCCGCGCCGCGCATGCTGCATACCGAGGGGCTACCTTGAGGC 237
DB 221 GlyThrProThrProAlaGlyValLeuSerThr-----ArgGlyProValSerArgGly 238
QY 238 CGTGT-----CGCACCTGTACAAACCC---TTGAGAGCTGGCGCG 276
DB 239 ArgGlyLeuLeuThrProArgAlaArgGlyValProProThrGlyTyrArgProProPro 258
QY 277 CCCCACCCCAATCCCGGCTATGCGGAGTAGTATCAAGAGCCAGTATGCAAT 336
DB 259 ProProProThrGlnGluThrTyrGlyGlyTyrAspTyrAspAspGly---TyrGlyThr 277
QY 337 AAATGCTACAGGGGTGTTAGCTGATACCGCTACGCGCCAGCCACCCTGCACTGCT 396
DB 278 AlaTyrAspGluGlnSerTyrAspSerTyrAspAsnSerTyrSerThrProAlaGlnSer 297
QY 397 GCTGCTACAGTACAGTACGAGAGGAGATTATGCTGCCGACCCCTAC 444
DB 298 AlaAlaAspTyrTyrAspTyrGlyHisGlyLeuSerGluAspAlaTyr 313

RESULT 14
US-09-442-100-8
; Sequence 8, Application US/09442100
; Patent No. 6359193
; GENERAL INFORMATION:
; APPLICANT: Xu, Tian
; APPLICANT: Tao, Wufan
; APPLICANT: Wang, Weiyl
; APPLICANT: Zhang, Sheng
; APPLICANT: Yu, Wan
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
; TITLE OF INVENTION: GENES AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442,100
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/411,111
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MISTOCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6523-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 980 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-442-100-8

Alignment Scores:
Pred. No.: 1.41 Length: 980
Score: 87.50 Matches: 50
Percent Similarity: 36.04% Conservative: 21
Best Local Similarity: 25.38% Mismatches: 81
Query Match: 7.52% Indels: 45
Gaps: 9

US-09-809-545a-1_COPY_535_1143 (1-609) x US-09-442-100-8 (1-980)

QY 76 CCCGACTTCTATGACGAGCGAGTCTGTGCGAGCGCAACGAGGATCTTCATG 135
DB 108 ProserhegiuglythnglylualaleuProserYrhsglnleuglyGlyAlaasn 127
QY 136 TACAGTGGCCCGACGTCATGTAATATCTTCGATGCTT-----GGCTTT 183
DB 128 TyrglunlglyProalalaleu-----gluglunetProAryglnlyLeuasphe 144
QY 184 CCAATCGCGCGCGCGACGTCGCGAGTCATACGAGGCGCTCAGCTTCGAGGCGGTGT 243
DB 145 LeuAheProgly-----AlaglyAlaglyThrhsglyAlaAlaAlaAlaAlaAla 162
QY 244 CGCAGCGGTACACACACCTTCAGAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 294
DB 163 Prolysglytyrserthr-----AlaValgluProserAlaAlaAlaAlaAlaAla 180
QY 295 --GCCATGCGCGAGAGTGTATCAAGACGAGTGTATGCG----- 333
DB 181 TyrglyAryglnlyslleuLeuSerlglnlserglYtyrglyValAlaArgseSer 200
QY 334 -----ATAAATGCTACAGAGGTGTACGCTGATACGCTGATACGCTGATACGCTG 378
DB 201 PheGlnAsnlystrProProAspAlaItyrserSerMetAlaAlaAlaAlaAlaAla 220
QY 379 -----CCACCG 426
DB 221 ProAlaSerleuthrPheProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 240
QY 427 TATGCTCCGACCGCTTACACACGAGCTGCTGCGCGCGCGCGCGCGCGCGCGCG 477
DB 241 AlaAlaAlaThrProProGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 260
QY 478 GTTGTGTCATGATGCTTTTGGCGCGCTGACCGATGCGCAAGACTAGAGCGCATG 537
DB 261 ThrGlyIuSerSerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 274
QY 538 GATGTGGGTCTCGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 588
DB 275 -----AasrleuAsnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 285

RESULT 15
US-08-552-142A-11
Sequence 11, Application US/08552142A
Patent No. 5695995
GENERAL INFORMATION:
APPLICANT: Weintraub, Harold M.
APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes
TITLE OF INVENTION: and Proteins
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen O'Connor Johnson KindnessP LLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/552,142A
FILING DATE: 02-NOV-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FPCR-1-8933
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-225-0709
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-552-142A-11

Alignment Scores:
Pred. No.: 1.16 Length: 379
Score: 87.00 Matches: 38
Percent Similarity: 33.52% Conservative: 23
Best Local Similarity: 20.88% Mismatches: 73
Query Match: 7.47% Indels: 48
Gaps: 6

US-09-809-545a-1_COPY_535_1143 (1-609) x US-08-552-142A-11 (1-379)

QY 40 TGGAAATTAATTCAGTGTGGCGCGGCTTACAGCGCGCATCTTATGACGAGCGGTG 99
DB 171 ThrAlaLeuSerlglnlyleuArgSerGlylyAsnArgProAspLeuValSerTyValGln 190
QY 100 CTGTGTGCG----- 108
DB 191 ThrleuCylysglyleuSerGlnProThrhAsnleuValAlaGlyCylyleuGlnleu 210
QY 109 -----CAGGCCACGAGGAGGATCTTCATGATGATGATGATGATGATGATGATG 147
DB 211 AsnSerArgAsnleuthrGlnGlnGlyArgAspAlaAlaAlaAlaAlaAlaAlaAla 230
QY 148 AGTTCATGTAATTAATCTTGCATGCTGCGCTTTCATATCGCGCGCGCGCGCGCG 198
DB 231 Glyly-----ProPheAlaMetAlaProtyrProtyrProtyrProtyrProtyr 247
QY 199 -----ACTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 243
DB 248 ArgThrValProGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 267
QY 244 CGC---ACCGGTACACACGCTTTCAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 297
DB 268 TyrcyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 287
QY 298 TATGCGGAGTGTATGACAGCGCGAGTGTATGCAATTAATGCTACAGGCGGTAC 357

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 12, 2003, 21:18:54 ; Search time 15 Seconds
(without alignments)
3424.218 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 1164

Sequence: 1 atgacatacaaaaggccgt.....acaaccgtttgtccacat 609

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 188354 segs, 42170167 residues
Total number of hits satisfying chosen parameters: 376708

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+n2p model -DEV=x1h
-Q=/cgn2_1/USFTO.spool/US09809545/runat_11032003_130947_12355/app_query.fasta_1.775
-DB=Published Applications_AA -QFMT=fastan -SUFIT=n2p.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62
-TRANS=human40.ccd -LIST=45 -DOCALIGN=200 -THR SCORE=pcr -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09809545 @CGN 1.1 9 @runat_11032003_130947_12355
-NCPU=6 -ICPU=3 -NO_XLPRX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	93.3	203	US-09-809-545A-2	Sequence 2, Appl1
2	624	53.6	330	US-09-794-591-2	Sequence 2, Appl1
3	98.5	8.5	1168	US-09-919-603-2	Sequence 2, Appl1
4	96	8.2	760	US-08-754-311B-2	Sequence 2, Appl1

5	95.5	8.2	168	10	US-09-864-761-40976	Sequence 40976, A
6	95.5	8.2	1172	10	US-09-919-770-4	Sequence 4, Appl1
7	95.5	8.2	1172	10	US-09-822-682-2	Sequence 2, Appl1
8	95	8.2	230	10	US-09-874-062-3	Sequence 3, Appl1
9	92	7.9	219	10	US-09-922-469-2	Sequence 2, Appl1
10	92	7.9	219	10	US-09-922-480-2	Sequence 2, Appl1
11	92	7.9	219	10	US-09-923-236-2	Sequence 2, Appl1
12	90.5	7.8	993	9	US-10-004-551-14	Sequence 14, Appl1
13	90.5	7.8	5179	9	US-10-025-380-1068	Sequence 1068, Ap
14	90.5	7.8	5179	10	US-09-922-217-1068	Sequence 1068, Ap
15	90.5	7.8	5179	10	US-09-833-263-1068	Sequence 1068, Ap
16	90.5	7.8	5179	10	US-09-764-877-1367	Sequence 1567, Ap
17	88.5	7.6	171	10	US-09-870-759-134	Sequence 134, App
18	88	7.5	830	9	US-09-870-759-134	Sequence 140, App
19	88	7.5	196	9	US-09-989-920-224	Sequence 224, App
20	87	7.5	650	10	US-09-784-358-10	Sequence 8, Appl1
21	87	7.5	724	10	US-09-784-358-8	Sequence 14, Appl1
22	87	7.5	724	10	US-09-784-358-14	Sequence 14, App
23	87	7.5	771	10	US-09-870-759-134	Sequence 140, App
24	87	7.5	830	9	US-09-870-759-140	Sequence 12, Appl1
25	87	7.5	845	10	US-09-784-358-12	Sequence 16, Appl1
26	87	7.5	845	10	US-09-784-358-16	Sequence 2, Appl1
27	87	7.5	1617	10	US-09-784-358-2	Sequence 2, Appl1
28	87	7.5	1631	10	US-10-024-450-2	Sequence 659, App
29	87	7.5	1706	9	US-09-925-297-659	Sequence 14, Appl1
30	86.5	7.4	171	10	US-10-235-674-14	Sequence 20, Appl1
31	86.5	7.4	262	9	US-09-263-689-14	Sequence 2, Appl1
32	86.5	7.4	262	9	US-09-263-689-14	Sequence 2, Appl1
33	86.5	7.4	874	9	US-10-020-215-2	Sequence 20, Appl1
34	86.5	7.4	1274	9	US-09-796-753-20	Sequence 2, Appl1
35	86	7.4	245	10	US-09-796-858-20	Sequence 20, Appl1
36	86	7.4	551	10	US-09-920-300A-1789	Sequence 1789, Ap
37	86	7.4	551	12	US-10-033-528-1789	Sequence 1789, Ap
38	86	7.4	1216	10	US-09-938-330-12	Sequence 12, Appl1
39	86	7.4	1222	10	US-09-938-330-8	Sequence 8, Appl1
40	86	7.4	1235	10	US-09-938-330-16	Sequence 16, Appl1
41	86	7.4	1235	10	US-09-938-330-20	Sequence 20, Appl1
42	86	7.4	1535	9	US-10-189-971-14	Sequence 14, Appl1
43	86	7.4	1570	9	US-10-189-971-12	Sequence 12, Appl1
44	85.5	7.3	310	9	US-10-084-094-12	Sequence 12, Appl1
45	85	7.3	231	10	US-09-864-761-37946	Sequence 37946, A

ALIGNMENTS

RESULT 1
US-09-809-545A-2
Sequence 2, Application US/09809545A
Patent No. US20020110804A1
GENERAL INFORMATION:
APPLICANT: Stanton, Lawrence W.
TITLE OF INVENTION: SECRETED FACTORS
FILE REFERENCE: SCIOS.017A
CURRENT APPLICATION NUMBER: US/09/809,545A
CURRENT FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 203
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-809-545A-2

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.65e-88	1086.00	100.00%	100.00%	93.30%	203	203	0	0	0	0

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-809-545A-2 (1-203)

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTG 60
DB 1 MetThrAnLysLysAlaValAsnProTyrThrAsnGlyTribLysLeuAsnProValVal 20
QY 61 GCGCGGCTCTACAGCCCGCACTTCTATGCAAGCAGGTCTGTTGGCCAGCCCAACCG 120
DB 21 GlyAlaValAltySerProAspPheTyrAlaGlyThrValLeuLeuGlnAlaAsnGln 40
QY 121 GAGGATCTTCATGATGACGTGGCCCGCACTTCACTGTATATCTTCGCAATGCTGSC 180
DB 41 GtuglySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
QY 181 TTTCCATATCCGCGCCGCACTGCTGAGCTGCAATCCAGGGGCTTCACCTTCAGAGCCGT 240
DB 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrAlaGlyAlaAlaSerLeuAlaValArg 80
QY 241 GGTCCACCGCTGTACACACCTTCAGAGCTGGGGCCCGCCCAATCCCGGCTAT 300
DB 81 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATTCGACAGGTTATCGCT 360
DB 101 GlyGlyValValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
QY 361 GCATACCGCTACGCCGCAAGCCACCCCTGCGCAGCTGCTGCTGCTACAGTACAGTACGGA 420
DB 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerLeuAspTyrGly 140
QY 421 CGAGTTTATGCTGCGGACCCCTACACACACACTTGTCTCCAGCCCGCCACTTACGGCGTT 480
DB 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
QY 481 GGTGCATGAATGCTTTGGCGCTTGACCGATGCCAAGTACAGGAGCATGCTGATGAT 540
DB 161 GlyAlaMetAsnAlaPheAlaPheLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
QY 541 GTGGGCTCTGTTCTTCTTCTTATGACGCTAGTATATACCAAGGGGATACACCGTTT 600
DB 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200
QY 601 GCTCCATAT 609
DB 201 AlaProTyr 203

RESULT 2
US-09-794-591-2
; Sequence 2, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulez, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-794-591-2

Alignment Scores:
Pred. No.: 2,996-47 Length: 330
Score: 624.00 Matches: 119
Percent Similarity: 89.78% Conservative: 4
Best Local Similarity: 86.86% Mismatches: 12
Query Match: 53.61% Indels: 2

DB: 10 Gaps: 0
US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-794-591-2 (1-330)
QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTG 60
DB 1 MetThrAnLysLysAlaValAsnProTyrThrAsnGlyTribLysLeuAsnProValVal 215
QY 61 GCGCGGCTCTACAGCCCGCACTTCTATGCAAGCAGGTCTGTTGGCCAGCCCAACCG 120
DB 21 GlyAlaValAltySerProGlnPheTyrAlaGlyThrValLeuLeuGlnAlaAsnGln 235
QY 121 GAGGATCTTCATGATGACGTGGCCCGCACTTCACTGTATATCTTCGCAATGCTGSC 180
DB 236 GtuglySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255
QY 181 TTTCCATATCCGCGCCGCACTGCTGAGCTGCAATCCAGGGGCTTCACCTTCAGAGCCGT 240
DB 256 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 275
QY 241 GGTCCACCGCTGTACACACCTTCAGAGCTGGGGCCCGCCCAATCCCGGCTAT 300
DB 276 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 295
QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATTCCTACAGGTTGTTACGT 360
DB 296 GlyGlyValValTyrProGlyTribLeuTyrCyAsArgHisLeu--TribTribLeuCySc 315
QY 361 GCATACCGCTACGCCGCAAGCCACCCCTGCGCAGCTGCTGCTGCTACAGTACAGTACGGA 407
DB 315 YHisProLeuArgProAlaTyrProCyHisCyAsArgCyLeuGln 330

RESULT 3
US-09-919-603-2
; Sequence 2, Application US/09919603
; Patent No. US20020137679A1
; GENERAL INFORMATION:
; APPLICANT: Lawler, John W.
; TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
; TITLE OF INVENTION: Chimeric Proteins
; FILE REFERENCE: 1440 1033-007
; CURRENT APPLICATION NUMBER: US/09/919,603
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: PCT/US00/02482
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/118,053
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1168
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-919-603-2

Alignment Scores:
Pred. No.: 0.905 Length: 1168
Score: 98.50 Matches: 50
Percent Similarity: 35.55% Conservative: 25
Best Local Similarity: 23.70% Mismatches: 66
Query Match: 8.46% Indels: 71
DB: 10 Gaps: 13

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-919-603-2 (1-1168)
QY 12 AAAGCCGCGTAACCCCTACACCAATGCTGGAATTAATTCAGTTGTGCGCGGTCTA 71
DB 380 GtuglyTribSerProTribAlaGlnTribThrGlnCySerValThrCySelySerGlyThr 399
QY 72 GAGCCCGGAGCT-----TCTATGACG 91
DB 400 GlnGlnArgGlyArgSerCyAspValThrSerAsnThrCySLeuGlyProSerIleGln 419

```

QY 92 GCACGCTGCTGT-----TGTCACAGCCCAACGAGGAGAT----- 127
Db 420 ThrArgAlaCysSerLeuSerLysCysAspThrArgIleArgGlnAspGlyLysTrpSer 439
QY 128 -----CTTCATATGTACAGTGGCCCACTTCATGTAT 160
Db 440 HisTrpSerProTrpSerSerCysSerValThrCysGlyValGlyAsnIle-ThrArgIle 459
QY 161 ATACTTGCATATGCTGGCTT---TCCATATCCGCGCCCACTGCTGACGTGCATACC 217
Db 459 eArgLeuCyAsnSerProValProGlnMetGlyGlyLysAsnGly----- 474
QY 218 GAGGGCTCAGCTTCGAGGCCGTGTCACCGCTGACACACCTTCAGAGCTGC--GG 274
Db 475 -LysGlySerGly-----ArgGlnThrLysAlaCysGlnGln 486
QY 275 CGCCGCCACCCCAATCCCGCTTATGCGGAGTAGTATCAAGAGCACTGATGCA 334
Db 486 yAlaProCysProIleAspGlyArgTrp-----SerProTrp-- 498
QY 335 ATAAATGCTACAGAGGTGTACGCTGATACCGCTACGCCAGCCACCCCTGCACATG 394
Db 499 -----SerProTrpSerAlaCysThrVal-----Th 507
QY 395 CTGCTGCTACAGTACAGTACGAGAGTGTATGCTGCCAGCCCTACACACACAC-- 452
Db 507 rCyAlaGlyGlyLysIleArgGlnArgThrArgValCysAsnSerProGlnProGlnTyrGln 527
QY 453 -----ACTTGC---TCCAGCCCCCAGCTTACGGCTTGGTCCATGATCTTTGGCC 502
Db 527 yGlyLysAlaCysValGlyAspValGlnGlnArgGlnMetCysAsnLysArgSerCysPr 547
QY 503 CCTGACCGATGCCAAGCTACGAGCCATGC 533
Db 547 oValAspGlyCysLeuSer---AsnProCys 556

```

RESULT 4

```

US-08-754-311B-2
; Sequence 2, Application US/08754311B
; Patent No. US20020004221A1
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Teet, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/754,311B
; FILING DATE: 21-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/195,152
; FILING DATE: 14-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249

```

```

; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-754-311B-2

```

Alignment Scores:

```

Pred. No.: 1.38 Length: 760
Score: 96.00 Matches: 42
Percent Similarity: 37.29% Conservative: 24
Best Local Similarity: 23.73% Mismatches: 61
Query Match: 8.25% Indels: 50
DB: Gaps: 6

```

US-09-809-545A-1_COPY_535_1143 (1-609) x US-08-754-311B-2 (1-760)

```

QY 124 GGATCTTCATGTAC-----AGTGGCCCACTTCACTT 156
Db 223 GlySerAsnLeuTyrGlyCysSerSerAlaSerAsnProLeuAspGlyValAlaValAla 242
QY 157 GTATATACCTTGTCAATGCTGCTTCATATATCCGCGCCACTGCTGCAGCTGCATAC 216
Db 243 ValAsnSerSerAlaVal-----AlaAlaAlaAlaAlaAlaValTyr 256
QY 217 CGAGGGGCTACCTTCGAGCGCGGTGTCACCGTGTACACACCTTCAGAGCTGCGGCG 276
Db 257 AspGlyLysHis-----AspTyrTyrTyrTyrAsnSerMetGlnGlnTyrThr 272
QY 277 CCCCCACCCCAATCCCGGCTATGCGGAGTAGTATCAAGACCAAGTATGGAAT 336
Db 273 ProProProPheTyrSerGlyTyrGlyThrProTyrAlaAlaAlaAlaAlaAlaArgGln 292
QY 337 AAATGCTACAGGCTGTACGCTGCATACCGCTACGCCACCCCTCCCTGCT 396
Db 293 AlaLysMetGlnProGlyAlaAlaAlaAlaAlaAlaTyrLeuThrProSerTyrAla 312
QY 397 GCTGCC----- 402
Db 313 AlaSerGlyAsnAsnAsnSerGlnLeuTyrSerSerProTyrAlaGlyTyrAsnAsnAsn 332
QY 403 -----TACAGTACAGTACGAGCAGATTTATGCTGCCGAC 438
Db 333 GlyGlnGlnAspTyrGlyGlyTyrTyrAsnGlnGlnTyrGlyAsnTyrTyrSerProAla 352
QY 439 CCTTACCAACACACACTTCTCCAGCCCC-----ACCTACGCGCTTGT-- 483
Db 353 AsnTyrSerProTyrAlaValAsnSerProSerSerSerAlaSerHisGlyHisGlyPhe 372
QY 484 GCCATGATGCTTTGGCCCTTGACCGATGCCAAGATAGAGCCATGCT 534
Db 373 HisValAlaAlaSerSerAsnLeuSerGlnSerProThrAspThrHisSer 389

```

RESULT 5

```

US-09-864-761-40976
; Sequence 40976, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26

```

```

PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,667
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 4917
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 40976
LENGTH: 168
TYPE: prt
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC011815.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
OTHER INFORMATION: EST HUMAN HIT: T70393.1, EVALUE 5.0e-17
OTHER INFORMATION: SWISSPROT HIT: P31794, EVALUE 2.1e-01
US-09-864-761-40976

```

Alignment Scores:					
Pred. No. :	1-12	length:	168		
Score:	95.50	Matches:	42		
Percent Similarity:	35.22%	Conservative:	14		
Best Local Similarity:	26.42%	Mismatches:	50		
Query Match:	8.20%	Indels:	53		
DB:	10	Gaps:	11		
 US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-864-761-40976 (1-168)					
Oy	15 GGCCTAAACCCCTACACCATGTGCTGGAAATTAAA-----TGCATTTGGGCC	65			
Dd	8 gLyrrgInPrroglrIgtTrpLcndlnlwnlvYvclYuglYlgYcyvAvglvs	27			
Oy	66 GGT-----CTACAGCCCCGAATTCTATAGCAGGCAAGCGTGCTGTGTTCAGGSCAA	116			
Dd	28 gLyserCyvleVleWlgnlnlseserCyvleWlYeurprnlsserCyvleWlYeuProdlasee	47			
Oy	117 CAGAGAGGATCTTCATGTATACAGTAgGCCCAATTCACCTTGATATACTTCTGCATGCC	176			
Dd	48 LyvaenThrvll-----CyseerLv	54			

```

QY 177 TGGCTTCGATATCGGCGCCGACCTGTGGAGCTGGCATACGAGGGC---TACCTTCG 233
      ::::
Db 55 -----AapLeuGlyGly-----CysSerCysrHrGInGInuSerArGAlaProAla 69
      ::::
QY 234 AGCGCGGTGTGCGACCGGTGACAA-----CACCTTCAGAGCTGC 272
      |||||
Db 70 CysSerTrpSerProArgThnGInGlyGlySerAlaSerHisAapLeuGlySerCys 89
      |||||
QY 273 -----GGCGGCCCCCAGCCGCCAACCCGGC----- 296
      |||||
Db 90 SerTyAlaTrpGlyAlaProThnArgGInLeuGlyArgSerTrpAlaProAlaCysPro 109
      |||||
QY 297 -----CTATGG-----CGAGTAGTGTATCAAGACCCAGTGTATGCGACATAAATT 341
      |||||
Db 110 ArgSerLeuTrpLeuCysGlyAlaHisSerMetGlyArgAlaSerTrpLeuGInProAla 129
      |||||
QY 342 GCTAAGAGGGGTGTAACGTCGATACCGGCTATCGGCCACGCCACCCGCTGACCTGTC 398
      ::::
Db 130 SerTrpGInTrp-----ProLeuGInMetValGInMetAlaHisCysCys 144
      ||||

```

```

RESULT 6
US-09-919-770-4
; Sequence 4, Application US/09919770
; Patent No. US20020048577A1
; GENERAL INFORMATION:
; APPLICANT: Bornstein, Paul
; APPLICANT: Kyriakides, Themis
; APPLICANT: Ratner, Buddy
; APPLICANT: Giachelli, Cecilia
; APPLICANT: Martinson, Laura
; APPLICANT: Scatena, Maria
; TITLE OF INVENTION: Methods and Devices to Modulate the Wound Response
; FILE REFERENCE: USFW117618
; CURRENT APPLICATION NUMBER: US/09/919,770
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,071
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1172
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-919-770-4

```

	Alignment Scores:			
	Pred. No.:	1.67	Length:	1172
	Score:	95.50	Matches:	50
	Percent Similarity:	35.55%	Conservative:	25
	Best Local Similarity:	23.70%	Mismatches:	66
	Query Match:	8.20%	Indels:	71
	DB:	10	Gaps:	13
US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-919-770-4 (1-1172)				
QY	12 AAAGGCGCTGAACCCCTTACACCATTGGCTGGAATTTAAATTCAGTTGTGGCGCGGTCTA	71		
Dd	382 GUGUgIyTrPserProTpRpaIagIutPrhCInCySseValThrCySgiSerGIyThr	40D		
QY	72 CAGCCCCGACT-----TCATGCAG	91		
Dd	402 GInGInArGlyArVatSerCySaSPvaIThSerAnThrCySeLugIyProserITlegIn	421		
QY	92 GCACGGTGCTCT-----TGTGCCAGCGCAACAGAGAGAT-----	127		
Dd	422 ThrAgAlaCySerLeuSerLyCySaSPThrrglIeArGInaPePgIyGLyTrPser	441		
QY	128 -----CTTCATGTACAGTGGCGCCCACTTCAGTTCACTGTAT	160		
Dd	442 HIsTrPserProTrPserSerCySeSValThrcySgiValGIyaEnIlE-ThrArgII	461		
QY	161 ATACTTCGCAATGCTGGCT---TCCATATCCGGCGCCCACTGCTGCAGCTGCATAC	217		

```
Db 461 eArygleuCyAsnSerProValProGlnMeGlyGlyLysAsnCyS----- 476
Qy 218 GAGGGGCTCACTTCGAGCGCGTGTGCACCGCTGTACACCTTCAGAGCTC---GG 274
Db 477 -LysGlySerGly-----ArgGlnThrLysAlaCySgIngl 488
Qy 275 CGCCCCCACCCTTCAGCGCGCTATGCGAGTAGTGATCAAGACCCAGTGTANGCA 334
Db 488 yAlaProCySProlleAspGlyArgTPr-----SerProTPr-- 500
Qy 335 ATAAATTGTACAGGGGTTCACCTACATCCGCTACGCCACCCAGCCCTGCCACTG 394
Db 501 -----SerProTPrSerAlaCySthrVal-----Th 509
Qy 395 CTGCTGCTACAGTACGAGTTAGTTCAGAGTTATGCTTCGCCACCTTACCACACAC-- 452
Db 509 rCySaAlaGlyGlyLleArgGlnArgThrArgValCySaAsnSerProGlnProGlnTyrGl 529
Qy 453 -----ACTTGC---TCCAGCCCCCACCCTACGCGCGTGTGCTGATGATGCTTTGCGC 502
Db 529 yGlyLysAlaCySaValGlyAspValGlnGlnArgGlnMeCySaSlnLysArgSerCySPr 549
Qy 503 CCTTGACCCGATGCCAAGACTAGAGCCATGC 533
Db 549 oValAspGlyCySLeuSer---AsnProCyS 558

RESULT 7
US-09-822-682-2
; Sequence 2, Application US/09822682
; Patent No. US20020119921A1
; GENERAL INFORMATION:
; APPLICANT: Detmar, Michael J.
; TITLE OF INVENTION: THROMBOSPONDIN-2 AND USES THEREOF
; FILE REFERENCE: 10287-051002
; CURRENT APPLICATION NUMBER: US/09/822,682
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/536,087
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/127,221
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1172
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-682-2

Alignment Scores:
Pred. No.: 1 67 Length: 1172
Score: 95.50 Matches: 50
Percent Similarity: 35.55% Conservative: 25
Best Local Similarity: 23.70% Mismatches: 66
Query Match: 8.20% Indels: 71
Gaps: 13

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-822-682-2 (1-1172)
Qy 12 AAAGCCGTGAACCCCTACCAATGCTGTGAAATTAATCAGTTGTGGCGCGTCTA 71
Db 382 GlnGlyTrpSerProTPrAlaGlnTrpThrGlnCySerValThrCySgInglThr 401
Qy 72 CAGCCCGGACT-----TCTATGAG 91
Db 402 GlnGlnArgGlyArgSerCySaSPValThrSerAnThrCySLeuGlyProSerIleGln 421
Qy 92 GCAGCGGCTGT-----TGTGCCAGGCCAACCCAGAGGAT----- 127
Db 422 ThrArgAlaCySerLeuSerLysCySaSPThrArgIleArgGlnAspGlyLysTPrSer 441
Qy 128 -----CTTCATGTACAGTGGCCAGTTCACTTGAT 160
```

```
Db 442 HisTPrSerProTPrSerSerCySserValThrCySglValGlyAsnIle-ThrArgIl 461
Qy 161 ATACTTGTGCATGACCTGGCTT---TCCATATCCGGCGCCACTGTGCGAGCATACC 217
Db 461 eArygleuCyAsnSerProValProGlnMeTgLyGlyLysAsnCyS----- 476
Qy 218 GAGGGGCTCACTTCGAGCGCGTGTGCACCGCTGTACACACCTTCAGAGCTC---GG 274
Db 477 -LysGlySerGly-----ArgGlnThrLysAlaCySgIngl 488
Qy 275 CGCCCCCACCCTTCAGCGCGCTATGCGAGTAGTGATCAAGACCCAGTGTANGCA 334
Db 488 yAlaProCySProlleAspGlyArgTPr-----SerProTPr-- 500
Qy 335 ATAAATTGTACAGGGGTTCACCTACATCCGCTACGCCACCCAGCCCTGCCACTG 394
Db 501 -----SerProTPrSerAlaCySthrVal-----Th 509
Qy 395 CTGCTGCTACAGTACGAGTTAGTTCAGAGTTATGCTTCGCCACCTTACCACACAC-- 452
Db 509 rCySaAlaGlyGlyLleArgGlnArgThrArgValCySaAsnSerProGlnProGlnTyrGl 529
Qy 453 -----ACTTGC---TCCAGCCCCCACCCTACGCGCGTGTGCTGATGATGCTTTGCGC 502
Db 529 yGlyLysAlaCySaValGlyAspValGlnGlnArgGlnMeCySaSlnLysArgSerCySPr 549
Qy 503 CCTTGACCCGATGCCAAGACTAGAGCCATGC 533
Db 549 oValAspGlyCySLeuSer---AsnProCyS 558

RESULT 8
US-09-874-062-3
; Sequence 3, Application US/09874062
; Patent No. US20020081607A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Four Disulfide Core Domain-Containing (FDCD) Polynucleotides,
; FILE REFERENCE: PT033P1
; CURRENT APPLICATION NUMBER: US/09/874,062
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32462
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/168,229
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-062-3

Alignment Scores:
Pred. No.: 1 32 Length: 230
Score: 95.00 Matches: 56
Percent Similarity: 32.27% Conservative: 15
Best Local Similarity: 25.45% Mismatches: 65
Query Match: 8.16% Indels: 86
Gaps: 13

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-874-062-3 (1-230)
Qy 54 AGTTGTGGGCGCGGTCTTACA-----GCCCGACTTATACAGGACGG 97
Db 14 SerCySglYglnGlyLeuGlyGlnGlnLysSerCySaSPArgProSerCySgInThrThr 33
Qy 98 TGTGCT-----TGTGCCAGGCCAACCCAGAGGATTTCCATGTACAGGCGCCCA 148
Db 34 CyCySaArgThrThrCySaSPArgProSer-----Cys----- 44
Qy 149 GTTCACTTGTATATACTTGTGCAATGCTGTTCATATCCGCGCGCCAGCTGCTGC-- 206
```

```
Db 45 -----|||:|||||-----CysIleSerSerCysCybArgPro-SerCysCysI1 56
Qy 207 -----AGCTGATACCG-----AGGGCTACCTTCGAG 235
Db 56 eSerSerCysCybIysPProSerCysCybLeuThrThrcyCybArgThrThrcyCybArg 76
Qy 236 GCCGTG-----GTGGACCGGTGTACAAACCTTCAGAGCTGCGCGCCGCCACCCCA 288
Db 76 gProSerSerCysIleSerSerCysCybArgPProSer-----88
Qy 289 ATCCCGGCTATGCGGAGTAGTATCAAGCCAGCTGTATGCATTAATTCGTACAG 348
Db 89 -----CysCysIleSerSerCysCybIysPProSerCysCybArg 101
Qy 349 GGTGTTACGCTGATACCG-----CTAGCCGCGACCCACCCCT 387
Db 101 gThrThrCys-CybArgPProSerCysCybIleSerSerCysCybArgPProSerCysCybI 121
Qy 388 GCCACTGCTGCTGCTACAGTAGTACAGTACGAGAGTTATTCGCGACCC-----440
Db 121 leSerSerCysCybIysPProSerCysCybArgThrThrcyCybArgPProSerCysCybC 140
Qy 441 -----CTACCACCAACACTTGTCTCCAGCCCGCCACCTACGCGCTTGTC 486
Db 140 yIleSerSerCysCybArgPProSerCysCybIleSerSerCysCybIysPProSerCysCybC 160
Qy 487 ATGAATGCTTTTGCGCCCTTGACCGATGCCAGACTAGGACGATGATGTGGGT 546
Db 160 yEgInThrThrcyCybArgPProSerCys-----Cys-----170
Qy 547 CTCGTTCTTTTCATTCGAGGCTATGATATACCAAGGGGATACACCGTTTTC 602
Db 171 -----IleSerSerCysTyArgPProGlnCysGlnPProSerCys 184

RESULT 9
US-09-922-469-2
; Sequence 2, Application US/09922469
; Patent No. US20020173027A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,469
; PRIOR FILING DATE: 2001-08-03
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-469-2

Alignment Scores:
Pred. No.: 2,4 Length: 219
Score: 92.00 Matches: 53
Percent Similarity: 33.91% Conservative: 6
Best Local Similarity: 30.46% Mismatches: 56
Query Match: 7.90% Indels: 59
DB: 9 Gaps: 10

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-922-469-2 (1-219)
Qy 24 CCCCTACACCAATGCTGGAATTAATCCAGTTGTGGGCGCGGTACAGCCCGCACTT 83
Db 33 ProlenHisPProSerLeuAnIleProTyArgIyle---ArgAnleuPProPProleu 51
Qy 84 CTATGAGGACGACGCTGTGTGTCAGAGGCAACGAGAGGATCTTCATGTACAGTGG 143
Db 52 TyTyTyArgPProValenThrValPProSerTyProGlyAnThrTyx-ThrAspThgl 71
```

```
Qy 144 CCCAGTTCACTGTATATATTCTGCAATGCTTCCTTTCCATATCCGGCCGCACTGC 203
Db 71 yLeuPProSerTyxProTrpIleuThrSerProGlyPheProTyx-----86
Qy 204 TGCAGCTGCAATACGAGGGGCTGACCTTCGAGC-----237
Db 87 -----ValTyx-----HisIleArgGlyPheProleuAlaThrGlnleuAnva 101
Qy 238 -----CGTGTTCGACACGCTGTAC-----AACACTTCAGAGCTGC 272
Db 101 lProProleuPProProArgIyPheProPheValPProProSerArgPhePheSerAlaI 121
Qy 273 GGGCCCCCA-----CCCCCAATCCGGCTTATGCGGAGTACTGTATCAAGCCAGT 326
Db 121 aAlaAlaPProAlaAlaPProIleAlaAla-----GluProAl 134
Qy 327 GTATGCAATTAATTCGTACAGGGGTGTACGTGATACCGCTACGCCGACCCACCC 386
Db 134 a-----AlaAlaAlaPProleuTh 140
Qy 387 TGCCACTGCTGCTGCTACAGTAGTACAGTACGAGAGTTATTCGCGACCCCTACCA 446
Db 140 rAlaThrProValAlaAlaIagIuPProAlaIagIyAlaPProValAlaAlaIagIuPro-----158
Qy 447 CCACACTTGTCTCCAGCCCGCCACCTACGCGCTTGTC 486
Db 159 -----AlaAlaIagIuAlaPro-----ValGlyAla 166

RESULT 10
US-09-922-480-2
; Sequence 2, Application US/09922480
; Patent No. US20020081701A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; PRIOR FILING DATE: 2001-08-03
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-480-2

Alignment Scores:
Pred. No.: 2,4 Length: 219
Score: 92.00 Matches: 53
Percent Similarity: 33.91% Conservative: 6
Best Local Similarity: 30.46% Mismatches: 56
Query Match: 7.90% Indels: 59
DB: 10 Gaps: 10

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-922-480-2 (1-219)
Qy 24 CCCCTACACCAATGCTGGAATTAATCCAGTTGTGGGCGCGGTACAGCCCGCACTT 83
Db 33 ProlenHisPProSerLeuAnIleProTyArgIyle---ArgAnleuPProPProleu 51
Qy 84 CTATGAGGACGACGCTGTGTGTCAGAGGCAACGAGAGGATCTTCATGTACAGTGG 143
Db 52 TyTyTyArgPProValenThrValPProSerTyxProGlyAnThrTyx-ThrAspThgl 71
Qy 144 CCCAGTTCACTGTATATATTCTGCAATGCTTCCTTTCCATATCCGGCCGCACTGC 203
Db 71 yLeuPProSerTyxProTrpIleuThrSerProGlyPheProTyx-----86
Qy 204 TGCAGCTGCAATACGAGGGGCTGACCTTCGAGC-----237
```

Db 87 -----ValTyr-----HisIleArgGlyPheProLeuAlaThrGlnLeuAsnVa 101
Qy 238 -----CGTGTCCGACCGGTGAC-----AACACCTTCAGAGCTGC 272
Db 101 lProProLeuProProArgGlyPheProPheValProProSerArgPhePheSerAlaAl 121
Qy 273 GGGCCCCCA-----CCCCCAATCCGGCCTTAGCGGAGTACTGATCAAGCCAGT 326
Db 121 aaIaIaIaProAlaIaIaProIleAlaIaIa-----GluProAl 134
Qy 327 GTATGCAATAAATTTGCTACAGGGGTGTACGCTGCATACCGCTACGCCCAACCC 386
Db 134 a-----AlaIaIaIaProLeuTh 140
Qy 387 TGCCACTGCTGCTCCTACAGTACGATGACGAGATTATGCTGCGACCCCTACCA 446
Db 140 rAlaThrProValAlaIaIaGluProAlaIaIaGlyAlaIaProValAlaIaIaGluPro----- 158
Qy 447 CCACACACTTGCTCCAGCCCCCACCCTACGGCGTTGTGCC 486
Db 159 -----AlaIaIaGluAlaPro-----ValGlyAla 166
RESULT 11
US-09-923-236-2
; Sequence 2, Application US/09923236
; Patent No. US20020090677A1
; GENERAL INFORMATION:
; APPLICANT: Shepard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; PRIORITY FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-236-2
Alignment Scores:
Pred. No.: 2.4 Length: 219
Score: 92.00 Matches: 53
Percent Similarity: 33.91% Conservative: 6
Best Local Similarity: 30.46% Mismatches: 56
Query Match: 7.90% Indels: 59
Gaps: 10
US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-923-236-2 (1-219)
Qy 24 CCCCTACCAATGCTGGAATTAATTCAGTTGTGGCCGCGGTCTACAGCCCCGACTT 83
Db 33 ProLeuHisProSerLeuAsnIleProTyrGlyIle---ArgAsnLeuProProLeu 51
Qy 84 CTATGACGACGCGGTGCTTTGTGCGACGCCAACCCAGAGGAGGATCTTCATGACAGTG 143
Db 52 TyrTyrArgProValaAsnThrValProSerTyrProGlyAsnThrTyr---ThraPThrG 71
Qy 144 CCCAGTTCACTTATATATCTTCTGCAATGCTGGCTTCATATCCGCGCGGCACAGC 203
Db 71 yLeuProSerTyrProTyrIleLeuThrSerProGlyPheProTyr----- 86
Qy 204 TGACGTGATATCCGAGGGGCTCACTTCAGGC----- 237
Db 87 -----ValTyr-----HisIleArgGlyPheProLeuAlaThrGlnLeuAsnVa 101
Qy 238 -----CGTGTCCGACCGGTGAC-----AACACCTTCAGAGCTGC 272
Db 101 lProProLeuProProArgGlyPheProPheValProProSerArgPhePheSerAlaAl 121

Qy 273 GGGCCCCCA-----CCCCCAATCCGGCCTTAGCGGAGTACTGATCAAGCCAGT 326
Db 121 aaIaIaIaProAlaIaIaProIleAlaIaIa-----GluProAl 134
Qy 327 GTATGCAATAAATTTGCTACAGGGGTGTACGCTGCATACCGCTACGCCCAACCC 386
Db 134 a-----AlaIaIaIaProLeuTh 140
Qy 387 TGCCACTGCTGCTCCTACAGTACGAGATTATGCTGCGACCCCTACCA 446
Db 140 rAlaThrProValAlaIaIaGluProAlaIaIaGlyAlaIaProValAlaIaIaGluPro----- 158
Qy 447 CCACACACTTGCTCCAGCCCCCACCCTACGGCGTTGTGCC 486
Db 159 -----AlaIaIaGluAlaPro-----ValGlyAla 166
RESULT 12
US-10-004-551-14
; Sequence 14, Application US/10004551
; Publication No. US20030004310A1
; GENERAL INFORMATION:
; APPLICANT: SHIMKETS, RICHARD A
; APPLICANT: FERNANDES, EUMA
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 15966-559
; CURRENT APPLICATION NUMBER: US/10/004,551
; PRIORITY FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 09/635,949
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 993
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: n 1755 can be A, G, C, or T
US-10-004-551-14
Alignment Scores:
Pred. No.: 4.44 Length: 993
Score: 90.50 Matches: 57
Percent Similarity: 38.27% Conservative: 36
Best Local Similarity: 23.46% Mismatches: 79
Query Match: 7.77% Indels: 72
Gaps: 12
US-09-809-545A-1_COPY_535_1143 (1-609) x US-10-004-551-14 (1-993)
Qy 34 AATGGCTGGAATTAATTCAC-----GTTGGGCGCGGTCTACAGCCCC 78
Db 718 AsnGlyTyrIlys---SerProSerGlnProGluLeuValHisGlyThrValThrTyr 736
Qy 79 GACTTCTATGACAGC-----ACGGTGTGTTGTCCAGC----- 113
Db 737 GlnCysTyrProGlyTyrGlnValaIaIaGlySerSerValLeuMetCysGlnTyrPAspLeu 756
Qy 114 -----CAACCAAGAGCG 125
Db 757 ThrTyrSerGlnAspLeuProSerCysGlnArgValThrSerCysHisAspProGlyAsp 776
Qy 126 ATCTTCAT-----GTACAGTGGCCCGGCTTC-----ACTTGATA 161
Db 777 ValGlnHisSerArgArgLeuIleSerSerProIysPheProValaIaIaThrValaGln 796
Qy 162 TACTTCTGCAATGCTGCTTCATATCCGCGGCACCTGCTGACAGTGCATATCCGAG 221
Db 797 TyrIleCysAspGlnGlyPheValaIaIaThrGlySerSerIleLeuThrCysHisAspArg 816
Qy 222 GGCCTACCTTCAGAGCGCGTGTGACCGT-----GTACCAACCTTCAG 266
Db 817 GlnAlaIaIaSerProIysTyrSerAspArgAlaIaIaProIysCysLeuLeuGlnLeuLys 836

```
QY 267 AGTCG-----GGCGCCCCCACCCTATCCGGCCTATGG 302
    |||
Db 837 ProCysHISglYleuSerAlaProGluasnGlyAlaArgSerProGluGlnLeuHis 856
QY 303 CGAGTATGATATCAAGCCAGCTATAGGCATTAATTGCTACAGGGTGG----- 353
    |||
Db 857 ProAlaGly-----AlaThrIleHisPheSerCysAlaProGlyTyrValLeuLys 873
QY 354 -----TTACGCTGCATACCGCTACCGCCAC-----CCC 386
    |||
Db 874 GlyAlaIaSerIleLeuCyValProGlyHisProSerHisIleTyrSerAspProPro 893
QY 387 TGCCACTGCTGCTGCTACAGTACAGTACGAGCAAGTTATGCTGCCAGCCCTACCA 446
    |||
Db 894 -IleCysArgAlaIaSerLeuAspGlyPhe-----TyrAsnSerArgSerLeuAs 910
QY 447 CCACACACTTGTCTCAAGCCCACTACAGCGCTGTGCGCATGATGCTTTGGCCCTT 506
    |||
Db 910 pValaIaLysAlaProAlaIaSerSerThrLeuAspAlaAlaHisIleAlaAlaAla 930
QY 507 GACCGATGCCAAGACTAGAGCAGCTGATGATGGGCTCGTTCTTTCTTCAATGCA 566
    |||
Db 930 ePhe-LeuProLeuValAlaMetValLeuValGlyGlyValTyrPheTyrPheSerA 950
QY 567 GGCTA 571
    |||
Db 950 rGleu 951

RESULT 13
US-10-004-551-16
; Sequence 16, Application US/10004551
; Publication No. US20030004310A1
; GENERAL INFORMATION:
; APPLICANT: SHIMKETS, RICHARD A
; APPLICANT: FERNANDES, ELMA
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 15966-559
; CURRENT APPLICATION NUMBER: US/10/004,551
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 09/635,949
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 16
; LENGTH: 994
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: n 1755 can be A, G, C, or T.
US-10-004-551-16

Alignment Scores:
Pred. No.: 4 45 Length: 994
Score: 90.50 Matches: 57
Percent Similarity: 38.27% Conservative: 36
Best Local Similarity: 23.46% Mismatches: 79
Query Match: 7.77% Indels: 72
DB: 9 Gaps: 12

US-09-809-545a-1_copy_535_1143 (1-609) x US-10-004-551-16 (1-994)
QY 34 AATGCTGGAATTAATCA-----GTTGTGGCGCGGTCTTACAGCCCC 78
    |||
Db 718 AanglyTrrPlyse---serProserGlnProGluLeuValHisIleGlyThrValAlaThrTyr 736
QY 79 GACTTATGCAAGC-----ACGCTGCTGTGGCCAGGC----- 113
    |||
Db 737 GlnCySerTyProGlyTyrGlnValValGlySerSerValLeuMetCyGlnTrrPleu 756
QY 114 -----CAACGAGGAGG 125
    |||
Db 757 ThrTrrSerGlnAspLeuProSerCyGlnArgValThrSerCyHisAspProGlyAsp 776
QY 126 ATCTTCAT-----GTACAGTGCGCCAGTTC-----ACTGTATA 161
```

```
Db 777 ValGlnHisSerArgAlaGlyLeuIleSerSerProLysPheProValGlyAlaThrValGln 796
    |||
QY 162 TACTTTCGATAGCCTGCTGCTTCCATACCGCGCCAGCTGCTCAGCTGCATGCAAGG 221
    |||
Db 797 TyrIleCysAspGlnGlyPheValLeuThrGlySerSerIleLeuThrCysHisAspArg 816
QY 222 GGTACACTTTCGAGCGCGTGTGCGACCGT-----GTACAAACACTTCAG 266
    |||
Db 817 GlnAlaGlySerProLysTrrPrrSerAspArgAlaProLysCysLeuLeuGlnGlnLeuLys 836
QY 267 AGTCG-----GGCGCCCCCACCCTATCCGGCCTATGG 302
    |||
Db 837 ProCysHISglYleuSerAlaProGluasnGlyAlaArgSerProGluLysGlnLeuHis 856
QY 303 CGAGTATGATATCAAGCCAGCTATAGGCATTAATTGCTACAGGGTGG----- 353
    |||
Db 857 ProAlaGly-----AlaThrIleHisPheSerCysAlaProGlyTyrValLeuLys 873
QY 354 -----TTACGCTGCATACCGCTACCGCCAC-----CCC 386
    |||
Db 874 GlyAlaIaSerIleLeuCyValProGlyHisProSerHisIleTyrSerAspProPro 893
QY 387 TGCCACTGCTGCTGCTACAGTACAGTACGAGCAAGTTATGCTGCCAGCCCTACCA 446
    |||
Db 894 -IleCysArgAlaIaSerLeuAspGlyPhe-----TyrAsnSerArgSerLeuAs 910
QY 447 CCACACACTTGTCTCAAGCCCACTACAGCGCTGTGCGCATGATGCTTTGGCCCTT 506
    |||
Db 910 pValaIaLysAlaProAlaIaSerSerThrLeuAspAlaAlaHisIleAlaAlaAla 930
QY 507 GACCGATGCCAAGACTAGAGCAGCTGATGATGGGCTCGTTCTTTCTTCAATGCA 566
    |||
Db 930 ePhe-LeuProLeuValAlaMetValLeuValGlyGlyValTyrPheTyrPheSerA 950
QY 567 GGCTA 571
    |||
Db 950 rGleu 951

RESULT 14
US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stoik, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedrick Thomas S.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: fastseq for windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068
```


Alignment Scores:
Pred. No.: 6.24 Length: 5179
Score: 90.50 Matches: 38
Percent Similarity: 33.83% Conservative: 7
Best Local Similarity: 28.57% Mismatches: 43
Query Match: 7.77% Indels: 45
DB: 9 Gaps: 6

US-09-809-545a-1_copy_535_1143 (1-609) x US-10-025-380-1068 (1-5179)

```
QY 141 TGGCCCGAGTTCATTTGATATCTTGCAGATGCT-----GGCTTTCATATCCGGC 194
DB 1390 TTPProleuetsplysCysIle-ThrThrProSerProProThrThrThrProSerPro-P 1409
QY 195 CGCCACTGTGCAGCTCATATACGAGGGGCTCACTTCGAGCCGGTGCACCGTGTGTA 254
DB 1409 roPro-----ThrThrThrThrThrLeuProProThrT 1420
QY 255 CAACACCTTCA-----GAGCTGGGGGCCCCCACCACCCCAATCCGGCTATGG 302
DB 1420 hrThrProSerProProThrThrThrThrThrThrProProThrThrThrProSerP 1440
QY 303 CGAGTGTGTATCAAGACGAGGTATGCAATTAATGCTACAGGTGTACG----- 358
DB 1440 roPro-----IleThrThrThrT 1446
QY 359 -----CTGCATACCGCTACGCCAGCCCGCCCTGCACCTGCTGCTG----- 400
DB 1446 hrThrProleuProThrThrThrProSerProProIleSerThrThrThrThrProProP 1466
QY 401 -----CTTACAGTGAAGTTACGAGAGGAGTTATGCTGCGAGCC 440
DB 1466 roThrThrThrProSerProProThrThrThrThrProSerProProThrThrThrProSerP 1486
QY 441 CTACACACACACACTTGCTCCAGCCCGCCACTAGC 475
DB 1486 roProThrThrThrThrThrThrThrProProProThr 1497
```

RESULT 15

US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US2002007641A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagner, Madeline Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Alignment Scores:

Pred. No.: 6.24 Length: 5179
Score: 90.50 Matches: 38
Percent Similarity: 33.83% Conservative: 7
Best Local Similarity: 28.57% Mismatches: 43

Query Match: 7.77% Indels: 45
DB: 10 Gaps: 6

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-922-217-1068 (1-5179)

```
QY 141 TGGCCCGAGTTCATTTGATATCTTGCAGATGCT-----GGCTTTCATATCCGGC 194
DB 1390 TTPProleuetsplysCysIle-ThrThrProSerProProThrThrThrProSerPro-P 1409
QY 195 CGCCACTGTGCAGCTCATATACGAGGGGCTCACTTCGAGCCGGTGCACCGTGTGTA 254
DB 1409 roPro-----ThrThrThrThrThrLeuProProThrT 1420
QY 255 CAACACCTTCA-----GAGCTGGGGGCCCCCACCACCCCAATCCGGCTATGG 302
DB 1420 hrThrProSerProProThrThrThrThrThrThrProProThrThrThrProSerP 1440
QY 303 CGAGTGTGTATCAAGACGAGGTATGCAATTAATGCTACAGGTGTACG----- 358
DB 1440 roPro-----IleThrThrThrT 1446
QY 359 -----CTGCATACCGCTACGCCAGCCCGCCCTGCACCTGCTGCTG----- 400
DB 1446 hrThrProleuProThrThrThrProSerProProIleSerThrThrThrThrProProP 1466
QY 401 -----CTTACAGTGAAGTTACGAGAGGAGTTATGCTGCGAGCC 440
DB 1466 roThrThrThrProSerProProThrThrThrThrProSerProProThrThrThrProSerP 1486
QY 441 CTACACACACACACTTGCTCCAGCCCGCCACTAGC 475
DB 1486 roProThrThrThrThrThrThrThrProProProThr 1497
```

Search completed: March 12, 2003, 22:22:08
Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2003, 22:52:53 ; Search time 1912 Seconds
(without alignments)
3089,890 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTRKAVNPYNGWKLNPV.....VLSSLQASIVGGYRFPAPY 203

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09809545/rnatc_07032003_153853_11454/app.query.fasta_1.391
-DB=GenEmbl -QFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=00 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09809545 @CGN 1 1 3745 @runat_07032003_153853_11454 -NCPU=6 -ICPU=3
-NO ALPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_ncg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_vl:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htg_hum:*
40: em_htg_mus:*
41: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1340	6	AX268800 Sequence
2	1063	97.9	2000	9	AF229057 Homo sapi
3	1061	97.7	1586	10	AF191501 Mus muscu
4	1024.5	94.3	1363	10	AF107204 Mus muscu
5	1021.5	94.1	3348	9	AF109106 Mus muscu
6	1020.5	94.0	2002	10	AB041596 Mus muscu
7	1009.5	93.0	1513	9	AK001027 Homo sapi
8	1009.5	93.0	2279	9	AF107203 Homo sapi
9	968	89.1	1475	9	AB060859 Macaca fa
10	955	87.9	2372	6	AR134676 Sequence
11	912	84.0	1547	9	AF094849 Homo sapi
12	526	48.4	1393	9	AY072786 Homo sapi
13	526	48.4	1535	9	BC025281 Homo sapi
14	526	48.4	1925	9	AF229058 Homo sapi
15	525	48.3	1134	10	AF387322 Mus muscu
16	525	48.3	1695	10	AF229055 Mus muscu
17	522	48.1	1538	9	BC013115 Homo sapi
18	519	47.8	1558	9	AK055213 Homo sapi
19	470.5	43.3	1721	10	BC027263 Homo sapi
20	467.5	43.0	1876	9	HSP38A20
21	456	42.0	959	10	AF229056 Mus muscu
22	347	32.0	1623	6	HSMB803066 Sequence
23	294	27.1	428	6	AX336745 Sequence
24	266.5	24.5	164538	9	AC005774 Homo sapi
25	262	24.1	159565	10	AL581075 Homo sapi
26	257	23.7	251	9	HSFOX09
27	253	23.3	209461	2	AC079359 Homo sapi
28	242.5	22.3	208632	2	AC073624 Homo sapi
29	231.5	21.3	2397	9	AK091695 Homo sapi
30	202	18.6	2971	9	HSFOX14
31	202	18.6	115860	2	AC120661 Homo sapi
32	193	17.8	596	9	HSAX33688 Homo sapi
33	193	17.8	617	9	HSAX35811 Homo sapi
34	193	17.8	653	9	HSAX35616 Homo sapi
35	193	17.8	675	9	HSAX34013 Homo sapi
36	167	15.4	203649	10	AL603843 Mouse DNA
37	164	15.1	108315	9	HS41P2
38	164	15.1	156820	2	AC123391 Homo sapi
39	145	13.4	73929	2	AC103207 Ratrus no
40	140	12.9	234	9	HSFOX13
41	133	12.2	374	9	HSFOX08
42	125	11.5	156549	2	AC079072 Homo sapi
43	123.5	11.4	7503	3	CCAZ51919 Cerialis
44	118.5	10.9	9704	10	MMU25652 Mus muscu
45	118.5	10.9	118496	2	AC123318 Ratrus no

RESULT 1

ALIGNMENTS

AX268800
LOCUS AX268800 1340 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE
ORGANISM Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS Stanton, L.W. and White, R.T.
TITLE Secreted factors
JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scios Inc. (US)
FEATURES
source location/Qualifiers
1..1340
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN
Alignment Scores:
Pred. No.: 1,6e-71 Length: 1340
Score: 1086.00 Matches: 203
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-09-809-545A-2 (1-203) x AX268800 (1-1340)
QY 1 MetThrAsnLysLeuValaValaAsnProTyrThrAsnGlyTTrpLysLeuAsnProValaVal 20
DB 535 ATGACTAATAAAGAGCCGCTGACCCCTTACACCAATGCTGGAAATTAATCCAGTTG 594
QY 21 G1AValaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyseGlnaLaangin 40
DB 595 GCGCGGCTTACAGCCCGCCGCTTCTATGACGACGCGTGTGTGCGCAGCCAG 654
QY 41 G1UGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
DB 655 GAGGGATCTTCCATGTACAGTGGCCCGAGTTCCTGTATATCTTCTGCAATGCTGGC 714
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
DB 715 TTTCATATCCGGCCGCGCACTGCTGACGCTGATACCGAGGGCTCACTTCAGGCCGT 774
QY 81 G1ArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
DB 775 GGTCCGACCGTGAACAACCTTCAAGAGTGGCGGCCCGCCCACTCCCGGCTAT 834
QY 101 G1GlyValaValTyrGlnGluProValTyrGlyAsnLysLeuGlnGlyGlyTyrAla 120
DB 835 GCGCGAGTGTATATCAAGAGCCAGTGTGGCAATAAATTCCTACAGGGGTGTTACGCT 894
QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
DB 895 GCATACCGCTACGCCCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
QY 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
DB 955 CGAGTTATGCTGCGACCCCTACACACACACCTGCTCCAGCCCGCCAGCTACGCGCTT 1014
QY 161 G1AlaAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
DB 1015 GGTGCATTAAGTCTTTCGCGCTTACCGAGTCCCAAGCTAGAGCCCTGCTGATGAT 1074
QY 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyValTyrAsnArgPhe 200
DB 1075 GTGGGTCTGTTCTTCTTCATTCAGGCTAGTATATACCAAGGGGATACACCGCTTTT 1134

QY 201 AlaProTyr 203
DB 1135 GCTCCATAT 1143
RESULT 2
AF229057
LOCUS AF229057 2000 bp mRNA linear PRI 01-FEB-2002
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
(HNRNP1) mRNA, complete cds.
ACCESSION AF229057
VERSION AF229057.1 GI:18461366
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Chen, W. and Winkelman, J.C.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, The Vontz Center for
Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
FEATURES
source location/Qualifiers
1..2000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
1..2000
/gene="HNRNP1"
256..1512
/gene="HNRNP1"
/note="RNA-binding protein"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
gamma"
/protein_id="AA171904.1"
/db_xref="GI:18461367"
/translation="MLASQGVLLHPYGVMIVPAPYPLGLIOGNOEAAAADPTMAOP
YASQFAPQNGIPAEYTAPEHPAPEYTGQTVPEHTLNLNPPAQTSEOSPADTSA
QVSGATQTDADA PTGQPOPTOSSENTENKSOEKLHVSNIPEPRPDRLRQMGOF
SKIDVEITFERSRSKSGRTYFENSADADARERKLTVEGRKIVNNATAMTN
KTVNPTNGKLPVVGAVTSPEFVETVLLCQANQSGSMYSAPELVYTSAMPF
PYPAATAAARGALRGRTVNTFPAAPPPPIPAYGVVVOEPLYGNKLQGGY
AAYVQPTPATAAVSDSYGRVYAADPVHHALAPATYGVGAMNAFALTDATRSH
ADGVGVSSIQASIVYRGVYRPAFY"
BASE COUNT 502 a 505 c 487 g 506 t
ORIGIN
Alignment Scores:
Pred. No.: 1,24e-69 Length: 2000
Score: 1063.00 Matches: 198
Percent Similarity: 98.52% Conservative: 2
Best Local Similarity: 97.54% Mismatches: 3
Query Match: 97.88% Indels: 0
DB: Gaps: 0
US-09-809-545A-2 (1-203) x AF229057 (1-2000)
QY 1 MetThrAsnLysLeuValaValaAsnProTyrThrAsnGlyTTrpLysLeuAsnProValaVal 20
DB 901 ATGACAAATAAAGAGCCCTCAACCTTATACAAATGCTGGAAATTAATCCAGTTG 960
QY 21 G1AValaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyseGlnaLaangin 40
DB 961 GGTGAGTCTACAGTCCCAATTCATGACGACGCGTCTTGTGCGACGCAACCG 1020
QY 41 G1UGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
DB 1021 GAGGATCTTCCATGTACAGTCCCGCAGTTCACTGTATATACCTTCTGCAATGCGAGC 1080
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80

```

Db 1081 TTCCCGATTCAGAGCCAGCCGCGGCGCTACCGAGGCGGCACCTCGAGGCCGC 1140
Qy 81 GlyArgThrValTyrAsnThrPheArgAlaAlaProProProProIleProAlaTyr 100
Db 1141 GGTGCACCGGTGTACACACTTCAGGCGCGCGGCCCGCCCGCATCCGGCCCTAC 1200
Qy 101 GlyGlyValValTyrGlnGluProValTyrGlyAsnIleuLeuGlnGlyTyrAla 120
Db 1201 GCGGAGTAGTGTACAGACCGCTGTATGGCAATAATTGCTGCAAGGCGCTTATCT 1260
Qy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 1261 GCATACCGCTACGCGCACCTTACCCCTCCACTGCGCTCAAGTACAGTACCGGA 1320
Qy 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
Db 1321 CGATTATATGCTGCCAGCCCTTACACACGACCTTGCCTCCAGCCCGCCACCTAC 1380
Qy 161 GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaIleTyrArgSerHisAlaAsp 180
Db 1381 GGTCGATGATGCTTTTGACACTTGTACATGCGCAAGACATAGAGCATGCTGATAT 1440
Qy 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPhe 200
Db 1441 GTGGGTCTCGTCTTCTTCAATGACAGGCTAGTATATACCGAGGAGATACACCGTTT 1500
Qy 201 AlaProTyr 203
Db 1501 GCTCCATAC 1509

RESULT 3
LOCUS AF191501 1586 bp mRNA linear ROD 01-MAR-2002
DEFINITION Mus musculus hexaribonucleotide binding protein 1 (Hrbp1) mRNA,
complete cds.
VERSION AF191501
AF191501.1 GI:19032413
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1586)
AUTHORS Chen, W. and Winkelman, J.C.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
FEATURES
source
1..1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/issue_type="skeletal muscle"
1..1586
/gene="Hrbp1"
166..1419
/gene="Hrbp1"
/codon_start=1
/product="hexaribonucleotide binding protein 1"
/db_xref="GI:19032413"
/translating="T"
translation="MLASQGVLLHSYGVPMIVPAAPYRPGIMCGNQNQAAAAADPTMAOP
YASQAPSPONGICPAEYTAAPHPAPAEYTGQTTVPDHLINLYPTQTHSEGSATDSAG
TVSGTATCTDAAATPDGQPTQPSSENTSKQPKRLHYSNIPFRFPDRLQMGQCS
KILVEYIIFNERSGSGFVTFENSADRDAREKLGHTVVEGRKIENVNATAVWTK
KTVNPYINGWKLNPVGAIVSPDYAGTVLLCQANQESSMYSPSSLVYSAWGP
YPAASAAAYRGALRGGRGTGYTLRAAAPPIPAYGGVYQEPYVGNLTLCGGA
AYRAYOPTPATAAAYSDSGRYVADPYHTLAPAPYGVGMNMAFALTDAKTRSHA
DDVGLIVSSLSLQASTYRGYNFAPY"
BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 1,34e-69 Length: 1586
Score: 1061.00 Matches: 198
Percent Similarity: 98.52% Conservative: 2
Best Local Similarity: 97.54% Mismatches: 3
Query Match: 97.70% Indels: 0
DB: 10 Gaps: 0
US-09-809-545a-2 (1-203) x AF191501 (1-1586)
Qy 1 MethThraenlyslYsAlaValAsnProTyrThrAsnGlyTyrIleuAsnProValVal 20
Db 808 ATGACAAATTAAGAAAGCTGTAAACCCCTACACCAATGCTGGAATTAATCCAGTTG 867
Qy 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaAsnGln 40
Db 868 GCGCGGTCTACAGCCCGGACCTTCTATGACAGGACGGGTGCTGTGTCAGAGCCAAACAG 927
Qy 41 GluGlySerSerMetTyrSerGlyProSerSerIleuValTyrThrSerAlaMetProGly 60
Db 928 GAGGATCTTCCATAGTACAGTACAGGCCCGGCTTCACTGTATATCTTGTCAATGCTGGC 987
Qy 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 988 TTCCCATATCCGGCGGCTCTGTCTGCACTCATACGAGGGGCTCACCTTCGAGCGCT 1047
Qy 81 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
Db 1048 GGTGCACCGGGTACACACACCTCAGGGCTGACGGCGCCCGCCCAATCCGGCTAT 1107
Qy 101 GlyGlyValValTyrGlnGluProValTyrGlyAsnIleuLeuGlnGlyTyrAla 120
Db 1108 GCGGAGTAGTGTATCAAGACCCAGTGTATGCAATAATTGCTACAGGGTGTATGCT 1167
Qy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 1168 GCGTACCGCTATGCCACGCCACCCCTGCCACTGCCCTGCTACAGTACAGTACCGGA 1227
Qy 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
Db 1228 CGAGTTATGCTGCCAGCCCTTACACACACACACTTGTCCAGCCCGCCACCTAC 1287
Qy 161 GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaIleTyrArgSerHisAlaAsp 180
Db 1288 GGTGCATGAATGCTTTGCCCTTGCACCATGCGAAGCTAGAGCCAGCTGATAT 1347
Qy 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPhe 200
Db 1348 GTGGGTCTCGTCTTCTTCAATGACAGGCTAGTATATACCGAGGAGATACACCGTTT 1407
Qy 201 AlaProTyr 203
Db 1408 GCTCCATAT 1416

RESULT 4
LOCUS AF107204 1363 bp mRNA linear ROD 23-JUN-2000
DEFINITION Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
VERSION AF107204
AF107204.1 GI:8671587
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1363)
AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.
TITLE A novel protein, A2BP, with RNA binding motif, binds to C-terminal
ataxin-2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1363)
AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics

```


DB: 9 Gaps: 1

US-09-809-545a-2 (1-203) x AF109106 (1-3348)

Qy 1 MethThraenlyblybAlaValaAsnProtyrThraSnglyTbplybLeuAsnProvalaVal 20
 |||||
 Db 899 ATGCAAAATAAAGAACCGTCAACCTTATACAAATGGCTGAAATTAATCCAGTTGTG 958

Qy 21 G1yAlaValTySerProAspPheTyAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
 |||||
 Db 959 GGTGACAGCTACAGTCCCGAATTCATGACAGGACGCTCTGTTGTCAGAGCCCAACAG 1018

Qy 41 G1uGlySerSerMetTySerGlyProSerSerLeuValTyThrSerAlaMetProGly 60
 |||||
 Db 1019 GAGGATCTTCATGATGACAGTGGCCCGGAGTTCACCTTGATATACTTCTGCAATGCCAGGC 1078

Qy 61 PheProTyrrProAlaAlaThraAlaAlaAlaTyArgGlyAlaHisLeuAlaGlyArg 80
 |||||
 Db 1079 TTCCCGATCCAGACAGCCAGCCGCGCGCTTACCGAGGGCCACCTCGAGGCGCGC 1138

Qy 81 G1yArgThraValTyAsnThraPheArgAlaAlaAlaProProProG11ProAlaTy 100
 |||||
 Db 1139 GGTGACAGCTGATACACACCTTCAGGCGCGCGCGCGCGCGCGCGCGCGCTAC 1198

Qy 101 G1yGlyValValTyArgGlyProValTyArgAlaLeuLeuGlyGlyTyAla 120
 |||||
 Db 1199 GCGGAGATGATGATACAGAGCTGTGTATGGCAATTAATGCTCAGAGGTGTATGCT 1258

Qy 121 AlaTyArgTyAlaGlnProThraProAlaThraAlaAlaTySerAspSer----- 138
 |||||
 Db 1259 GCATATCCGCTACGCCACCTTACCCCTGCACTGCGCTGCTACAGACAG- AAATCA 1317

Qy 139 -----TyrGlyArgVal 142
 |||||
 Db 1318 GTTGCTCTTCGTCAGACAGATGAATTTCTTGAACACTCTGACGTTACGAGCACTT 1377

Qy 143 TyAlaAlaAspProTyrrThraHisThraLeuAlaProAlaProThrTyGlyValGlyAla 162
 |||||
 Db 1378 TATGTCGCGACCCCTTACACACGACCTTGTCTCAGGCCCGCCACTTACGGGTGTGCTCC 1437

Qy 163 MetAsnAlaPheAlaProLeuThraAlaValTyArgSerHisAlaAspAsyValGly 182
 |||||
 Db 1438 ATGATGCTTTTGACCTTTGATGATGCAAGACTAGAGCAAGCTGATGATGTGGGT 1497

Qy 183 LeuValLeuSerSerLeuGlnAlaSerIleTyArgGlyGlyTyAsnArgPheAlaPro 202
 |||||
 Db 1498 CTGCTCTTCTCTACATGACGCTAGTATATACGAGGGGAGTACACCGTTTGTCTCA 1557

Qy 203 Tyr 203
 |||||
 Db 1558 TAC 1560

RESULT 6
 AB041596
 LOCUS Mus musculus brain cDNA, clone MNCD-3035, similar to Homo sapiens
 DEFINITION cDNA FLJ10165, clone HEMBA1003591.
 AB041596 2002 bp mRNA linear ROD 30-JUN-2000
 AB041596.1 GI:7670455
 VERSION 1.1
 KEYWORDS full insert sequence
 SOURCE Mus musculus (strain: C57BL) adult female cDNA to mRNA,
 clone 11b: Sugano mouse brain mncb clone: MNCD-3035.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCES
 1 (sites)
 Osada, N., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and
 Hashimoto, K.
 TITLE isolation of full-length cDNA clones from mouse brain cDNA library
 made by oligo-capping method
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2002)
 AUTHORS Hashimoto, K., Osada, N., Kusuda, J. and Sugano, S.

TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of
 Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: hashiku@nii.go.jp, URL: http://www.nii.go.jp/yoken/genebank/,
 Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 URL: http://www.nii.go.jp/yoken/genebank/
 COMMENT
 Lab Name: Sugano mouse brain mncb
 Lab host: TOP10
 Vector: pME18s-FL3
 1st strand cDNA was primed with an oligo (dT) primer
 [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
 using specific 5' and 3' primers and amplified by PCR. The PCR
 product was digested with SfiI and size selection was performed to
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
 into distinct DraII sites of pME18s-FL3. XhoI sites just outside
 the DraII sites can be used to isolate the cDNA insert. Library
 was constructed by Sugano et al. (University of Tokyo, Institute of
 Medical Science). Custom primer used for sequencing (5' end primer
 [CTTGTGCTTAAAGCTGCG]; 3' end primer
 [CGACCTGACGCTCGACAC]).
 A part of this sequence is reported in AU067167.
 Location/Qualifiers
 source
 1..2002
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="MNCD-3035"
 /sex="female"
 /clone_11b="Sugano mouse brain mncb"
 /dev_stage="adult"
 637..1827
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA95079.1"
 /db_xref="GI:7670456"
 /translation="MNCERQLRGNDGAAAAPTMAQPVASQAFAPPONGIPAEYAP
 HPHAPAYTGQTVPGHTLNLVPTQTHSOSADTSQTSVGTATQDDAPDGGQ
 TOPSENEKSOPRLHLSNI PRFRPDRLDMGCGKLDVEI IENEGSKGFVY
 TPENSADADRAREKHGTVEGRKIEVNNAITARYMTKVNPTNCKLNPVYAVY
 SPDPYACTVILCOANDEGSSMSYGRPSLVYTSAMPGPYPAATTAAYRAHLRGR
 TVNTERFAAPPPPIPAYGVVVDGEGVADYIGVIAVYVYAOPTPAALAYSYGR
 VVADPYHTHTLAPTYGVGAMNNAFALTDKTRGHADVLVLSLQAS1YRGYNR
 FAYY"
 BASE COUNT 518 a 589 c 516 g 379 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,69e-66 Length: 2002
 Score: 1020.50 Matches: 192
 Percent Similarity: 96.06% Conservative: 3
 Best Local Similarity: 94.58% Mismatches: 7
 Query Match: 93.97% Indels: 1
 DB: 10 Gaps: 1

US-09-809-545a-2 (1-203) x AB041596 (1-2002)

Qy 1 MethThraenlyblybAlaValaAsnProtyrThraSnglyTbplybLeuAsnProvalaVal 20
 |||||
 Db 1219 ATGCAAAATAAAGAACCTGTAACCCCTTACACCAATGCTGAAATTAATCCAGTTGTG 1278

Qy 21 G1yAlaValTySerProAspPheTyAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
 |||||
 Db 1279 GCGCGGCTACAGCCCGGACTTCTATGACAGGACGCGTCTGTTGTCAGAGCCCAACAG 1338

Qy 41 G1uGlySerSerMetTySerGlyProSerSerLeuValTyThrSerAlaMetProGly 60
 |||||
 Db 1339 GAGGATCTTCATGATGACAGTGGCCCGGAGTTCACCTTGATATACTTCTGCAATGCCAGGC 1398

Qy 61 PheProTyrrProAlaAlaThraAlaAlaAlaTyArgGlyAlaHisLeuAlaGlyArg 80
 |||||
 Db 1399 TTCCCGATCCAGACAGCCAGCCGCGCTTACCGAGGGCCACCTCGAGGCGCGC 1458

QY 81 G1yAgtThrValTyrAsnThrPheArgAlaAlaProProProProIleProAlaTyr 100
 Db 1459 GGTCCACCGGTGTACAAACCTTCAGGGCTGCGAGCCCCCGCCCAATCCGGGCTAT 1518
 QY 101 G1yG1yValValTyrGlnGluProValTyrG1yAsnLysLeuGlnGlnG1yG1yTyrAla 120
 Db 1519 GCGCGTGTGTGTACCAAGATGGATTTATGGGCGAGAC--ATTATATGCTGATTATGCT 1575
 QY 121 A1ATyArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrG1y 140
 Db 1576 GCGTACCGGTATGCCAGCCACCTTCGACCTGCGCGCTCCTACAGTACACTTCGGA 1635
 QY 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrG1yVal 160
 Db 1636 CGAGTTTATGCTGCCAGCCCTACCAACACACTTGTCTCGAGCCCCCACTACGGCGCTT 1695
 QY 161 G1yAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp 180
 Db 1696 GGTCCATGAATGCTTTTCCGCTTGACCGAAGCTAGGGGCAATGCTGATGAT 1755
 QY 181 ValG1yLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnG1yG1yTyrAsnArgPhe 200
 Db 1756 GTGGTCTGCTCTTCTTCTTCATTGCGAGCTAGTATACCGAGGGGATACAACTTTT 1815
 QY 201 AlaProTyr 203
 Db 1816 GCTCCATAT 1824
 RESULT 7
 AK001027
 LOCUS Homo sapiens CDNA FLJ10165 f18, clone HEMBA1003591, weakly similar
 DEFINITION to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.
 ACCESSION AK001027
 VERSION AK001027.1 GI:7022045
 KEYWORDS oligo capping, f18 (full insert sequence).
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head CDNA to mRNA, clone lib:HEMBA1 clone:HEMBA1003591.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Isogai,T., Oea,T., Hayaishi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y., Nimomiya,K. and Iwayanagi,T.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1513)
 AUTHORS Isogai,T. and Otsuki,T.
 JOURNAL Direct Submission
 COMMENT Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
 FEATURES
 source
 location/Qualifiers
 1..1513
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1003591"
 /cissue_type="whole embryo, mainly head"
 /clone_lib="HEMBA1"
 /def_stage="embryo, 10 weeks"
 /note="cloning vector: pME18SFL3"

CDS
 221..1414
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91472.1"
 /db_xref="GI:7022046"
 /translation="MNCERROLRNGQMAAAPDTPMAOPYASQAPRPNQIPAEYTA
 PHPAPEYTGQTVPEHTLNPAPQTHSESPADTSQATISRTQMDADAPDGP
 QTPSENMENSOPEKRLVSNIPERFRPDRLQMFQGGKILDEVEIFNERGKGF
 VTPENSADADAREKLGHTVEGKIENVNATARTMTKTVNPYTNQMLNPVGA
 YSPFVACTVLLCOANQSGSMYSAPSLVYTSAMPFPYPAATAAAYRGAHLRG
 RTVNTPEPAAPPPPIPAYGVVODGVDGADIGVGAAYVAOPTPATTAAYSDSYG
 RYVADPPIHHAALAPPTYGVANMAFALUTPAKRSHADVDVILSSIQASIRGYN
 RPAFY"
 BASE COUNT 354 a 418 c 395 g 346 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 8e-66 Length: 1513
 Score: 1009.50 Matches: 190
 Percent Similarity: 95.57% Conservative: 4
 Best Local Similarity: 93.60% Mismatches: 8
 Query Match: 92.96% Indels: 1
 DB: 9 Gaps: 1
 US-09-809-545A-2 (1-203) x AK001027 (1-1513)
 QY 1 MetThrAsnLysLeuAlaValAsnProTyrThrAsnG1yTyrLysLeuAsnProValVal 20
 Db 806 ATGACAAATAAAAAGCCGTCAACCTTATACAAATGCTGGAAATGAAATTCAGTTGG 865
 QY 21 G1yAlaValTyrSerProAspPheTyrAlaG1yThrValLeuLeuLysGlnAlaAsnGln 40
 Db 866 GGTGAGCTCTACAGTCCCAATTTATGACGAGCGCTCTGTCGACGAGCCAAACG 925
 QY 41 G1yG1ySerSerMetTyrSerG1yProSerSerLeuValTyrThrSerAlaMetProG1y 60
 Db 926 GAGGAGATCTTCATGATGACGTGCCCCCACTTATATATCTTCTCCATGCCAGGC 985
 QY 926 GAGGAGATCTTCATGATGACGTGCCCCCACTTATATATCTTCTCCATGCCAGGC 985
 QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgG1yAlaHisLeuArgG1yArg 80
 Db 986 TTCCGCTATCCAGCAGCCACCGCGCGCGCTACCGAGGGGCGCACTGCGAGGCGCC 1045
 QY 81 G1yAgtThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
 Db 1046 GGTCCACCGGTGTACAAACCTTCAGGGCTGCGAGCCCCCGCCCAATCCGGGCTAC 1105
 QY 101 G1yG1yValValTyrGlnGluProValTyrG1yAsnLysLeuGlnGlnG1yG1yTyrAla 120
 Db 1106 GCGCGTGTGTGTACCAAGATGGATTTATGGGCGAGAC--ATTATATGCTGATTATGCT 1162
 QY 121 A1ATyArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrG1y 140
 Db 1163 GCATACCGGTATGCCAGCCACCTTCGACCTGCGCGCTCCTACAGTACGTTACGGA 1222
 QY 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrG1yVal 160
 Db 1223 CGAGTTTATGCTGCCAGCCCTACCAACACTTGTCTCGAGCCCCCACTACGGCGCTT 1282
 QY 161 G1yAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp 180
 Db 1283 GGTCCATGAATGCTTTTCCGCTTGACCTGAGTGAAGCTAGGAGCCATGCTGATGAT 1342
 QY 181 ValG1yLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnG1yG1yTyrAsnArgPhe 200
 Db 1343 GTGGTCTGCTCTTCTTCTTCATTGCGAGCTAGTATACCGAGGGGATACAACTTTT 1402
 QY 201 AlaProTyr 203
 Db 1403 GCTCCATAC 1411
 RESULT 8
 AF107203
 LOCUS AF107203 2279 bp mRNA linear PRI 23-JUN-2000


```

DEFINITION Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.
ACCESSION AF107203
VERSION AF107203.1 GI:8671585
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 2279)
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE A novel protein, A2BP, with RNA binding motif binds to C-terminal
ataxin-2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2279)
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA
FEATURES
source
1..2279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Similar to Homo sapiens cDNA clone HSP38A20"
1..2279
/gene="A2BP"
/locus="1047..2180"
/gene="A2BP"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AAF78291.1"
/db_xref="GI:8671585"
/translation="MAQPVASAPORPONGIPAEYTAHPHPAPAEYTGOTTPEHTLN
LYPPAQHRSQSPADTSQAQVSGRATDPAAPDGPQPSPTNKSKQPRKLNHS
NIPREFRDPDKQFGQFKLIDVEILFNERSGSGFGEFVENSADDRAREKLGIV
VEGRKLEVNNAATARVMTNKTVNPTVETGKMLNVPVAGVSPFVAGTVLLQANQES
SVYSAPSLVYTSMPGFPPYPAATAAAYAGAHVIRGRTVNTFRRAADPPPIPAYG
GVYODPFYADIDYGYAAVRAVQPTPATVAAVSDSYGRVYAADPYHHALAPATYGV
GANNAPALITDAKTRSHADVGLVSLQASIVRGYNRFAP"
misc_feature
1341..1358
/gene="A2BP"
/note="Region: RNA binding motif RNP-2"
1452..1472
/gene="A2BP"
/note="Region: RNA binding motif RNP-1"
BASE COUNT 497 a 712 c 636 g 434 t
ORIGIN
Alignment Scores:
Pred. No.: 1,26e-65 Length: 2279
Score: 1009.50 Matches: 190
Percent Similarity: 95.57% Conservative: 4
Best Local Similarity: 93.60% Mismatches: 8
Query Match: 92.96% Indels: 1
Gaps: 1
US-09-809-545a-2 (1-203) x AF107203 (1-2279)
QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTyrPLeuAsnProValVal 20
Db 1572 ATGACAAATAAAAGACCGTCAACCTTATACAAATGCTGGAATTGCAATCCAGTTGG 1631
QY 21 GAlaValAlaLysSerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
Db 1632 GGTGCAGCTCAAGTCCCGCAATTCATGACGACGGTCTGTGTGTCAGGCGCAACGAG 1691
QY 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 1692 GAGGATTTTCATATACAGTCCCGCACTTGTATATACCTTCGCAATGCCAGGC 1751
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80

```

```

Db 1752 TTCCCGTATCCAGACGCCCGCGCGCGCTACCGAGGGGGCGACCTGCCAGCGCC 1811
QY 81 GAlaArgThrValTyrAsnThrPheArgAlaAlaAlaProProProPoiLeProAlaTyr 100
Db 1812 GGTGCACCGGTGTACACACCTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 1871
QY 101 GAlaGlyValValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
Db 1872 GCGCGTGTGTTCACAGATGAGATTTATGTCAGAC--ATTATGCGGTATATGCT 1928
QY 121 AAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 1929 GCATACCGCTACGCCCGACCTACCCCTGCACCTCGCTGCTACAGAGATACCGGA 1988
QY 141 ArgAlaTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
Db 1989 CGAGTTATATGCTGCCGACCCCTACACACGACATTTGTCAGGCCCGCACCTAGCGGCTT 2048
QY 161 GAlaAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
Db 2049 GGTCGCATGAAATGCTTTGCACTTGACATGATCCAAAGACTAGAGCCATGCTGATGAT 2108
QY 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerLeTyrGlnGlyGlyTyrAsnArgPhe 200
Db 2109 GTGGGTCTCGTTCTTCTTCATTCGACGCTAGATATACGAGGGGGATACACCGTTT 2168
QY 201 AAlaProTyr 203
Db 2169 GCTCCATTC 2177
RESULT 9
AB060859 1475 bp mRNA linear PRI 13-JUN-2001
LOCUS AB060859
DEFINITION Macaca fascicularis brain cDNA clone: Ctra-11594, full insert
sequence.
ACCESSION AB060859.1 GI:13874510
VERSION AB060859.1
KEYWORDS oligo capping; fib (full insert sequence).
SOURCE Macaca fascicularis adult male temporal lobe right cDNA to mRNA,
clone 11b; macaque brain cDNA library Ctra clone: Ctra-11594.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.
REFERENCE 1 (sites)
AUTHORS Oseada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
TITLE Isolation of full-length cDNA clones from macaque brain cDNA
libraries
JOURNAL Unpublished
AUTHORS Hashimoto,K., Oseada,N., Hida,M., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181
Lab host: TOP10
Vector: pME18S-FU3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTTGG)
R. Site2: DraIII (CACCATGG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGTCGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FU3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by oligo-capping method
(Sugano et al., Institute of Medical Science, University of
Tokyo).
Custom primer used for sequencing

```

(5' end primer [CTTCTGCTCTTAAAGCTGCG] ;
3' end primer [CGACCTGAGCTCGAGCACA]).
Location/Qualifiers
1. 1475
/organism="Macaca fascicularis"
/db_xref="taxon:9641"
/clone="Qcra-11594"
/sex="male"
/tissue_type="temporal lobe right"
/clone_lib="macaque brain cDNA library Qcra"
/dev_stage="adult"
90. 1220
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB46877.1"
/db_xref="GI:13874511"

CDS

BASE COUNT 389 a 414 c 363 g 309 t
ORIGIN
Alignment Scores:
Pred. No.: 8,926-63 Length: 1475
Score: 968.00 Matches: 189
Percent Similarity: 87.33% Conservative: 4
Best Local Similarity: 85.52% Mismatches: 9
Query Match: 89.13% Indels: 20
DB: 9 Gaps: 2
US-09-809-545A-2 (1-203) x AB060859 (1-1475)
QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal 20
Db 681 ATGACAAATATAAAGACCGTCAACCTTATACAAATGCGTGAATTTGAATCCAGTTGTG 740
QY 21 GtTtAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 741 GGTGGGCTTACAGTCCGGAATTTCTATGACGACGCGCTCTGTGTGCCAGGCCAACCG 800
QY 41 GtGtLysSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 801 GAGGAGTCTTCATGATGACGTGCCCGCCAGTTCCTGTATATCTTCTGCAATGCCAGGC 860
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyValHisLeuArgGlyArg 80
Db 861 TTCGGGTATCCAGACGCCCGCGCGCGCTTACCGAGGGCGCACCTGCGAGGCCCGC 920
QY 81 GtTtArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
Db 921 GCCCGACCGGTGACAAACCTTCAGGGCGCGAGCCCCCGCGCATCCCGGCTAT 980
QY 101 GtGtLysValValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
Db 981 GCGCGTGTCTTACCAAGATGATTTATGTGTGACAGC---ATTATGTGTGTTATGCT 1037
QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSer----- 138
Db 1038 GCATACCGGTACGCCAGCTACCCCTGCCAGCGCGCTCTTACAGTGACAG-AAATCA 1096
QY 139 -----TyrGtLysArgVal 142
Db 1097 GTTCGCTTCTGTCAGCAGATGAATTTCTTGAACACCTTGCAGTTACGAGCAGTT 1156
QY 143 TtTtAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGtLysValGlyAla 162
Db 1157 TATGTGCCGACCCCTTACCAACGACCTTGTCTCCAGCCCCCTTACGCGGTGTGCTCC 1216
QY 163 MetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspValGly 182

Db 1217 ATGAATGCTTTTGCACCTTTCATGATGCAAGACTTACGACCTGATGATGTTGGT 1276
QY 183 LeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPheAlaPro 202
Db 1277 CTCGCTTCTTCTTCAATTCAGCGCTAGTATATACGAGGGGATACACCGTTTCTCCA 1336
QY 203 Tyr 203
Db 1337 TAC 1339

RESULT 10

ARI34676 2372 bp DNA linear PAT 16-MAY-2001
LOCUS ARI34676
DEFINITION Sequence 1 from patent US 6194171.
ACCESSION ARI34676
VERSION ARI34676.1 GI:14123581
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 2372)
AUTHORS Pulst, S.M. and Shibata, H.
TITLE Nucleic acids encoding ataxin-2 binding proteins
JOURNAL Patent: US 6194171-A 1 27-FEB-2001;
FEATURES
source location/Qualifiers
1. 2372
/organism="unknown"

BASE COUNT 548 a 726 c 646 g 452 t
ORIGIN
Alignment Scores:
Pred. No.: 1,386-61 Length: 2372
Score: 955.00 Matches: 189
Percent Similarity: 86.94% Conservative: 4
Best Local Similarity: 85.14% Mismatches: 9
Query Match: 87.94% Indels: 21
DB: 6 Gaps: 2
US-09-809-545A-2 (1-203) x ARI34676 (1-2372)

QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal 20
Db 1572 ATGACAAATATAAAGACCGTCAACCTTATACAAATGCGTGAATTTGAATCCAGTTGTG 1631
QY 21 GtTtAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 1632 GGTGAGTCTTACAGTCCGGAATTTCTATGACGACGCGCTCTGTGTGCCAGGCCAACCG 1691
QY 41 GtGtLysSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 1692 GAGGAGTCTTCATGATGACGTGCCCGCCAGTTCCTGTATATCTTCTGCAATGCCAGGC 1751
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyValHisLeuArgGlyArg 80
Db 1752 TTCGGGTATCCAGACGCCCGCGCGCGCTTACCGAGGGCGCACCTGCGAGGCCCGC 1811
QY 81 GtTtArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
Db 1812 GGTGCGACCGGTGACAAACCTTCAGGGCGCGAGCCCCCGCGCATCCCGGCTAT 1871
QY 101 GtGtLysValValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
Db 1872 GCGCGTGTCTTACCAAGATGATTTATGTGTGACAGC---ATTATGTGTGTTATGCT 1928
QY 120 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSer----- 138
Db 1929 TGCATACCGCTACGCCAGCTTACCCCTGCCAGCGCGTCTTACAGTGACAG-AAATC 1987
QY 139 -----TyrGtLysArgVal 142
Db 1988 AGTTCGCTTCTGTCAGCAGATGAATTTCTTGTAAACCTCTGCAAGTTACGAGAGAGT 2047

OY	142	ITyrTAlAAspprCoIyRhISHiThrlenuAlaProAlaPcOTrTyRGlyValGlyAl	162
Db	2048	TtATCTGCAGACCCCTACACCAGCACTTGCTCCAGCCCCCACCATTGCGGTGGTC	2107
OY	162	aMeASnAlaPheAlaProLeuThrAspAlaYSThrArgSerHisAlaAspValGl	182
Db	2108	CATGAATGTTTGGACCTTTGACGTGATGCCAAGACTGAGACCATTGCTGATGATGTGG	2167
OY	182	yLeuValLeuSerSerLeuGlnAlaSerIleTyrgInglyGlyTYraanaYGpheAlaPr	202
Db	2168	TCTCGTCTCTTCTCATTCAGCAGCTAGTAATAACGAGGGGATACAACCGTTTGCTCC	2227
OY	202	cTyr 203	
Db	2228	ATAC 2231	
RESULT 11			
AF094849		1547 bp	mRNA linear PRI 01-MAR-2002
LOCUS	AF094849		
DEFINITION	Homo sapiens hexaribonucleotide binding protein 1 isoform alpha		
ACCESSION	(HRNBP1) mRNA, complete cds.		
VERSION	AF094849		
KEYWORDS	AF094849.1 GI:19032365		
SOURCE	.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1547)		
JOURNAL	Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppee,B. and		
REFERENCE	Winkelmann,J.C.		
AUTHORS	Molecular cloning and chromosomal localization of a human brain,		
TITLE	heart and skeletal muscle specific RNA binding protein gene		
JOURNAL	homologous to fox-1 in Caenorhabditis elegans		
REFERENCE	unpublished		
AUTHORS	2 (bases 1 to 1547)		
TITLE	Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppee,B. and		
JOURNAL	Winkelmann,J.C.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (24-SEP-1998) Internal Medicine/Hematology-Oncology,		
TITLE	University of Cincinnati College of Medicine, 231 Bethesda Ave.,		
JOURNAL	Cincinnati, OH 45267-0508, USA		
FEATURES			
source	Location/Qualifiers		
gene	1..1547		
cds	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="3"		
	/map="16p13.3"		
	1..1547		
	/gene="HRNBP1"		
	254..1432		
	/name="HRNBP1"		
	/note="similar to Caenorhabditis elegans fox-1"		
	/codon_start=1		
	/product="hexaribonucleotide binding protein 1 isoform		
	alpha"		
	/protein_id="AA183405.1"		
	/db_xref="GI:19032366"		
	/translation="MLASQGLHPHYGVPMVPAAPYLPGILIGNOEAAAADPTMAOP		
	YASAQFARPONGIRAEYTAAPHRAPEYTGCTYTPEHTNLYPQAOTHSESPDPTS		
	QTSGATOTDDAATDGQPOTQPSENTENSSQPRHLVSNIPFRFPDLRFGGQ		
	KTLIDVELIFMERKSGKGFTFENSADABARELHCTVEGRKILENNATAMVN		
	KKTVAYPTNGKLNKPVGAVVSPERYACTVLICAMOGSSMSAPSLLVTSAMPGF		
	PYPATATAAARGAHLRGRGRVTVVTFPAALPPPIPAYGGVVQEPYGNKLLOGG		
	AAAYRAQTPPATDAAYSXGRVVAADPYHHALAAPPYVGMAASTYRGYNRFAY		
BASE COUNT	380 a 448 c 374 g 345 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,276-58	length:	1547
Score:	912.00	Matches:	172
Percent Similarity:	85.71%	Conservative:	2

Query Match:	84.73%	Mismatches:	3
DB:	9	Indels:	26
		Gaps:	1
US-09-809-545A-2 (1-203) x AF094849 (1-1547)			
QY	1 MetTTrAsnLysLysAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal	20	
DB	899 ATGACAAATATAAAMACCGTCAACCTTATATCAATATGAGCTGGAAATTAATCAATTG	958	
QY	21 GAlaValAlTYrSerProAspPheTYrAlaGlyTTrValLeuLeuCYGAlaAsnGln	40	
DB	959 GGTGCAGCTTACAGTCCCGAATTCTATGACAGCAACGCTCTGTTGTGTCAGGCCAACAG	1018	
QY	41 GluGlySerSerMetTYrSerGlyProSerSerLeuValTYrThrSerAlaMetProGly	60	
DB	1019 GAGGATCTTCCATGTACAGTCCCGCCAGTTCACTTGATATCTTCTGCAATGCAAGC	1078	
QY	61 PheProTYrProAlaAlaThrAlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArg	80	
DB	1079 TTCCCGTATCCAGCAACCGCCGCGCGCTTACCAAGGGGCGCACCCTGGAGGCGCG	1138	
QY	81 GAlaGTrnValTYrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTYr	100	
DB	1139 GGTGCCACCGTGTACAAACACTTCAGGGCCGGCGGCCGCCGCCGATCCGGCTTAC	1198	
QY	101 GlYglYValAlaTYrGlnGluProValTYrGlyAsnLysLeuGlnGlyTYrAla	120	
DB	1199 GCGGAGTGTGTATCAAGAGCGCTGTATGGCAATTAATGCTCAGAGGGTATGCT	1258	
QY	121 AlaTYrArgTYrAlaGlnProThrProAlaThrAlaAlaAlaTYrSerAspSerTYrGly	140	
DB	1259 GGATACCGGTACGCCACCTTACCCCTCCACTGGCGCTTACAGTACAGTACGGA	1318	
QY	141 ArgValAlTYrAlaAlaAspProTYrHisHisThrLeuAlaProAlaProThrTYrGlyVal	160	
DB	1319 CAGAGTTATGCTGCCGAGACCCCTTACCAACGACACTTGCTCCAGCCCCCACTACGGGCTT	1378	
QY	161 GlYAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp	180	
DB	1379 GGTGCCATG-----	1387	
QY	181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTYrGlnGlyTYrAsnArgPhe	200	
DB	1388 -----CTAGTATATATCCAGGGGGATACAAACGGTTT	1420	
QY	201 AlaProTYr 203		
DB	1421 GCTCCATAC 1429		
RESULT 12			
LOCUS	AY072786	1393 bp	mRNA linear PRI 22-MAR-2002
DEFINITION	Homo sapiens RNA binding motif protein 9 mRNA, complete cds.		
ACCESSION	AY072786		
VERSION	AY072786.1		GI:19584571
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 1393)		
AUTHORS	Norris,J.D., Fan,D., Sheik,A. and McDonnell,D.P.		
TITLE	A negative coregulator for the human ER		
JOURNAL	Mol. Endocrinol. 16 (3), 459-468 (2002)		
MEDLINE	21864496		
PUBMED	11875103		
REFERENCE	2 (bases 1 to 1393)		
AUTHORS	Norris,J.D., Fan,D. and McDonnell,D.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-JUN-2002) Pharmacology and Cancer Biology, Duke		
	University Medical Center, Research Drive, LSCC Building, Rm C264,		
	Durham, NC 27710, USA		

```

FEATURES
    source
        Location/Qualifiers
            1..1393
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="22"
                /map="22q13.1"
                /cell_line="HeLa"
            106..1278
                /note="RBM9; splice variant; RPA; repressor of tamoxifen
                transcriptional activity"
                /codon_start=1
                /product="RNA binding motif protein 9"
                /protein_id="AA067150.1"
                /db_xref="GI:19584572"
                /translation="MONEPLTPGNGFPARDSDGNOEPTTPDAMVOPFTTIPPPPP
                ONGIPTEYGVPHTDYAGOTGEHNLTYGSDHGEQSSNSPSTONGSLTTEGCAOTD
                GQSGTOSSENSSEKSTPKRLHVSNI PRPRPDRLRQMGFGKILDYELLIFENERGSK
                GGFRTFENSADADKARKLHGTVEGRKLEVNNTARVMNKVTPYANGMKLSPIVYGFPPTAAT
                VGAIVGPELVYASSFOADVSLGNDAAVPLSGRGGINITIPILSLPLVVGFPPTAAT
                AAAPFGAHLRGGRVVGAVRAVPTPAIPAYGVVYQDGFAGADLYGVAAVRAQPA
                TATATATAAAAAAASDYGRTVADPYHALAPAAISYGVAVASLYRGYSRFABY"
            442..711
                /note="Region: RNA recognition motif"
                /misc_feature
                    /note="Region: RNA recognition motif"
BASE COUNT      361 a      376 c      379 g      277 t
ORIGIN
Alignment Scores:
Pred. No.:      3.25e-30      length:      1393
Score:          526.00      Matches:      116
Percent Similarity: 59.55%      Conservative: 15
Best Local Similarity: 52.73%      Mismatches: 43
Query Match:     48.43%      Indels:      46
DB:              9      Gaps:      8

US-09-809-545A-2 (1-203) x AY072786 (1-1393)
QY      1 MetTrrAenLYeAlaValaAenProTYrThraenGlyTTrlyLeuAenProValaI 20
Db      703 ATGACCAATTAAGAGATGCTACACCATATGCAAATGTTGGAAATTAAGCCAGTAGTT 762
QY      21 GYAAlaValTYrSerProAenPheTYrAlaGlyThrValLeuLeuCy6GlnAlaAen--- 39
Db      763 GGAGCGTATATGCTCCGAGATTATGACGATCCAGCTTTCACAGCATGTGTCCCTA 822
QY      40 ---GlnGluGlySerSerMetTYrSerGlyProSerSer-----Leu 52
Db      822 GGCATGATGACGACGACATGCCCTCATCAGGAAGGGGGATCAACAATTACCTCTTA 882
QY      53 ValTYrThraSerAlaMetProGlyPheProTYrPro---AlaAlaThraAlaAlaAla 71
Db      883 ATCAGTCTCCCTTATAGTCTCTGCTTCCCTTACCTCACTGCACGCCACGCGCGCT 942
QY      72 TYrArgGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 91
Db      943 TTCAGAGGAGCCCATTTGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 999
QY      92 AlaProProProProLeuProAlaTYrGlyValValTYrGlnGluProValaTYrGly 111
Db      1000 GACCTCCCAACAGCCATCCCGCTATCCAGGTGTGTTTACAGAGAGGATTTACGGT 1059
QY      112 AenLYeLeuLeuGlnGlyTYrAlaAlaTYrArgTYrAlaGlnProThraProAlaThr 131
Db      1060 GCTGAC---CTCATGTGTGATATGACGCTTACAGATATGACAGCTGCTACTGCAACC 1116
QY      132 -----AlaAlaAlaTYrSerAenSerTYrGlyArgValaTYr 143
Db      1117 GCAGCCACCGCTGCTGACGCGCTGACGCGCTTACAGTACGCGCTTATGCGAGGGTATC 1176
QY      144 AlaAlaAenProTYrThraSerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 163
Db      1177 ACAGCGGAGCCCTAC---CATGCCCTTGCCTCCCGCTACGTATGAGTGTGCGCTGTG 1233
QY      164 AenAlaPheAlaPheLeuThraPheAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 183

```

```

Db      1233 -----
QY      184 ValLeuSerSerLeuGlnAlaSerIleTYrGlnGlyTYrAsnArgPheAlaProTYr 203
Db      1234 -----GCGAGTTTATACGAGGTGCTACAGCCGATTTGCCCTTAC 1275

RESULT 13
LOCUS      BC025281
DEFINITION Homo sapiens, RNA binding motif protein 9, clone MGC:39192
ACCESSION BC025281
VERSION   BC025281.1 GI:19264129
KEYWORDS   MGC.
SOURCE      Homo sapiens.
ORGANISM   Homo sapiens.
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 1555)
REFERENCE   Strausberg, R.
            Direct Submision
            Submitted (05-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
REMARK      Contact: MGC help desk
            Email: cga@bbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Sueanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Leticia Hsiao, Martin Krzywinski, Neta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 42 Row: d Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657503.

FEATURES
    source
        Location/Qualifiers
            1..1555
                /organism="Homo sapiens"
                /db_xref="LocusID:23543"
                /db_xref="taxon:9606"
                /clone="MGC:39192 IMAGE:4555865"
                /tissue_type="uterus, leiomyosarcoma"
                /clone_id="NIH_MGC_46"
                /lab_host="DH10B-R"
                /note="Vector: pOTB7"
            189..1331
                /codon_start=1
                /product="RNA binding motif protein 9"
                /protein_id="AAH25281.1"
                /db_xref="GI:19264130"
                /translation="MEKKMYTOGNOEPTTPDAMVOPFTTIPPPPPONGIPTEYGV
                PHTDYAGOTGEHNLTYGSDHGEQSSNSPSTONGSLTTEGCAOTDGOQSGTOSSE
                NSBKSTPKRLHVSNI PRPRPDRLRQMGFGKILDYELLIFENERGSKGFFRTFENS
                ADADARKLHGTVEGRKLEVNNTARVMNKVTPYANGMKLSPIVYGFPPTAATTAARGAHLR
                AASFOADVSLGNDAAVPLSGRGGINITIPILSLPLVVGFPPTAATTAARGAHLR
                GRGRVVGAVRAVPTPAIPAYGVVYQDGFAGADLYGVAAVRAQPAATATATATAAAA
                AAAAAVDYGRVYADPYHALAPAAISYGVAVASLYRGYSRFABY"
BASE COUNT      469 a      371 c      383 g      332 t
ORIGIN

```

Alignment Scores:

Pred. No.: 3,68e-30 Length: 1555
 Score: 526.00 Matches: 116
 Percent Similarity: 59.55% Conservative: 15
 Best Local Similarity: 52.73% Mismatches: 43
 Query Match: 48.43% Indels: 46
 DB: 9 Gaps: 8

US-09-809-545A-2 (1-203) x BC025281 (1-1555)

Qy 1 MetThrAsnLysValAlaAsnProTyrThrAsnGlyTTPDysLeuAsnProValVal 20
 Db 756 ATGACCAATAAGAAAGATGTCACACCATATGCACAAATGGTTGAAATTAAAGCCACTAGTT 815
 Qy 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsn--- 39
 Db 816 GGAGCTGTAATAGTCCCGAGATTATGACAGATCCAGCTTCAAGCAGATGTCTCCCTA 875
 Qy 40 ---GlnGlySerSerMetTyrSerGlyProSerSer-----Leu 52
 Db 876 GGCAATATGACGACGTCGCCCTATCAGGAAGAGGGGTATCAACACTTACATTCCTTTA 935
 Qy 53 ValTyrThrSerAlaMetProGlyPheProTyrPro---AlaAlaThrAlaAlaAla 71
 Db 936 ATCAGTCTCCCTTGTGTTCTGCTTCCCTTACCTACTGACGACGACGACGACGCT 995
 Qy 72 TyrArgGlyAlaHisLeuArgGlyValArgThrValTyrAsnThrPheArgAlaAla 91
 Db 996 TTCAGAGAGCCCATTTAAGGGGCGAGGCGGACAGATATGTTGTCAGTCCGA---GCG 1052
 Qy 92 AlaProProProProlleProAlaTyrGlyValValTyrGlnGluProValTyrGly 111
 Db 1053 GTACTCTCAACAGCATCCCGCCCTATCCAGGTGTGTTTACAGAGGAGGATTTACGGT 1112
 Qy 112 AsnLysLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
 Db 1113 GCTGAC---CTCTATGTGTGATATGACGCTCAGATATGACAGCCTGCTCAACAC 1169
 Qy 132 -----AlaAlaAlaTyrSerAspSerTyrGlyValArgValTyr 143
 Db 1170 GCAGCCACCGCTGTCGACCGCGCTGACGATGACGATGAGGTTATGCGAGGTATAC 1229
 Qy 144 AlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMet 163
 Db 1230 ACAGCCGACCCCTAC---CATGCCCTTGCCCTCCGCTAGCTATGAGCTTGCCCTGTG 1286
 Qy 164 AsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeu 183
 Db 1286 ----- 1286
 Qy 184 ValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPheAlaProTyr 203
 Db 1287 -----GCGAGTTTATACGAGGTGCTACAGCGCATTTGCCCTTAC 1328
 RESULT 14
 AF229058 1925 bp mRNA linear PRI 01-FEB-2002
 LOCUS AF229058 Homo sapiens hexaribonucleotide binding protein 2 (HRNP2) mRNA,
 DEFINITION complete cds.
 ACCESSION AF229058
 VERSION AF229058.1 GI:18461368
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 1925)
 TITLE Direct Submission
 AUTHORS Chen, W. and Winkelman, J. C.
 JOURNAL Submitted (28-JUN-2000) Internal Medicine/Hematology-Oncology,
 University of Cincinnati College of Medicine, The Vontz Center for
 Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA

FEATURES
source

Location/Qualifiers
 1..1925
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 1..1925
 /gene="HRNP2"
 171..1313
 /gene="HRNP2"
 /note="RNA-binding protein; alternatively spliced RBM9"

CDS

/codon_start=1
 /product="hexaribonucleotide binding protein 2"
 /protein_id="AA171905.1"
 /db_xref="GI:18461369"
 /translation="MEKKKVVTOGNOEPTTPDPAVQPTTIPPPPPONGIPREYGV
 PHODVAGOTGHNLTLYGSTOAHGEOSNSPTSONSLTTEGAOTDGOOSOTGSE
 NSSSKSPKRLHYSNIPFRPDDLRMPQFGKLLDVEIIFNRSKGRGFPYFNS
 ADDRAERKLGTVBGRKTEVNNAIKRVTNKMTVPYANGKLSFVGVAVPELY
 AASSFQADVSLGNDAAVPLSGRGINTYIPLSLPLVPGFPPTAAATTAAPRAHLR
 GGRITVGAIVAVPPTAIPAVPGVYODGFGADLYXGVAAYRAOQATATATMAAA
 AAAAYSDGYGRVYTAADPYHALAPASVGVAVASLYRGYSRFAPY"

BASE COUNT 562 a 436 c 467 g 460 t
 ORIGIN

Alignment Scores:

Pred. No.: 4.67e-30 Length: 1925
 Score: 526.00 Matches: 116
 Percent Similarity: 59.55% Conservative: 15
 Best Local Similarity: 52.73% Mismatches: 43
 Query Match: 48.43% Indels: 46
 DB: 9 Gaps: 8

US-09-809-545A-2 (1-203) x AF229058 (1-1925)

Qy 1 MetThrAsnLysValAlaAsnProTyrThrAsnGlyTTPDysLeuAsnProValVal 20
 Db 738 ATGACCAATAAGAAAGATGTCACACCATATGCACAAATGGTTGAAATTAAAGCCACTAGTT 797
 Qy 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsn--- 39
 Db 798 GGAGCTGTAATAGTCCCGAGATTATGACAGATCCAGCTTCAAGCAGATGTCTCCCTA 857
 Qy 40 ---GlnGlySerSerMetTyrSerGlyProSerSer-----Leu 52
 Db 858 GGCAATATGACGACGTCGCCCTATCAGGAAGAGGGGTATCAACACTTACATTCCTTTA 917
 Qy 53 ValTyrThrSerAlaMetProGlyPheProTyrPro---AlaAlaThrAlaAlaAla 71
 Db 918 ATCAGTCTCCCTTGTGTTCTGCTTCCCTTACCTACTGACGACGACGACGCT 977
 Qy 72 TyrArgGlyAlaHisLeuArgGlyValArgThrValTyrAsnThrPheArgAlaAla 91
 Db 978 TTCAGAGAGCCCATTTAAGGGGCGAGGCGGACAGATATGTTGTCAGTCCGA---GCG 1034
 Qy 92 AlaProProProProlleProAlaTyrGlyValValTyrGlnGluProValTyrGly 111
 Db 1035 GTACTCTCAACAGCATCCCGCCCTATCCAGGTGTGTTTACAGAGGAGGATTTACGGT 1094
 Qy 112 AsnLysLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
 Db 1095 GCTGAC---CTCTATGTGTGATATGACGCTCAGATATGACAGCCTGCTACTGCAAC 1151
 Qy 132 -----AlaAlaAlaTyrSerAspSerTyrGlyValArgValTyr 143
 Db 1152 GCAGCCACCGCTGTCGACCGCGCTGACGCTTACATGACGTTATGCGAGGTATAC 1211
 Qy 144 AlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMet 163
 Db 1212 ACAGCCGACCCCTAC---CATGCCCTTGCCCTCCGCTAGCTATGAGCTTGCCCTGTG 1268
 Qy 164 AsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeu 183
 Db 1268 ----- 1268

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 15, 2003, 16:22:47 ; Search time 225 Seconds
(without alignments)
2031.803 Million cell updates/sec

Title: US-09-809-545A-2
Perfect score: 1086
Sequence: 1 MTNKKAVNPYNGMKLNPNV.....VLSLSQASLYGYGYNREAPY 203

Scoring table:

BLOSUM62			
Xgapop 10.0 , Xgapext 0.5			
Fgapop 6.0 , Fgapext 7.0			
Delop 6.0 , Delext 7.0			

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```

-MODEL=frame+.p2n.model -DEV=xlp
-Q=/cgnr_1/USPTO.spool/US09809545/runat_07032003_153852_11445/app_query.fasta-1.391
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blcsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09809545 -GCN1_1_1396 @runat_07032003_153852_11445 -NCPY=6 -ICPY=3
-NO_XLPHY -NO_MMAR -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRESHOLD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database :

```

N_Geneseq_101002.*
1: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
24: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1340	AA594693	Rat secreted facto
2	1009.5	93.0	1513	AAH13824	Human cDNA sequenc
3	1009.5	93.0	1800	AAK52245	Human polynucleoti
4	955	87.9	2372	AAH07075	cDNA encoding huma
5	617	56.8	607	AAH09205	Human cDNA clone (
6	606.5	55.8	481	AA593633	DNA encoding novel
7	605	55.7	539	ABK34530	Human cDNA for nov
8	534.5	49.2	567	AA566245	DNA encoding novel
9	534.5	49.2	939	AA161034	Human polynucleoti
10	534.5	49.2	939	AA161035	Human polynucleoti
11	534	49.2	918	AA159249	Human polynucleoti
12	534	49.2	1011	AA159248	Human polynucleoti
13	511	47.1	1252	AA590510	DNA encoding novel
14	467.5	43.0	1506	AA246827	Human RNA binding
15	405.5	37.3	2118	AA593634	DNA encoding novel
16	368.5	33.9	578	AA590506	DNA encoding novel
17	342	31.5	406	AA593632	DNA encoding novel
18	294	27.1	428	ABL68917	Kidney cancer rela
19	258	23.8	1164	AAK53229	Human polynucleoti
20	239	22.0	327	AAK54001	Murine transcripti
21	231.5	21.3	413	AAK04077	Human secreted pro
22	218.5	20.1	407	AA186721	Human polynucleoti
23	198	18.2	316	AAK05035	Human brain expres
24	198	18.2	3189	ABK34602	Human cDNA for nov
25	129	11.9	1104	ABL24461	Drosophila melanog
26	116	10.7	9192	ABL92117	Human Tumour Endot
27	116	10.7	9483	AA544690	Human full-length
28	116	10.7	11657	ABV23446	Human prostate exp
29	116	10.7	11657	ABV29303	Human prostate exp
30	110.5	10.2	660	ABL07512	Drosophila melanog
31	110.5	10.2	2660	ABL07513	Drosophila melanog
32	110	10.1	109519	AA580693	Microsomopora DNA
33	108.5	10.0	6930	AA158020	Human polynucleoti
34	108.5	10.0	7272	AA159806	Human polynucleoti
35	107	9.9	895	ABN98713	Arabidopsis thalia
36	103.5	9.5	2834	AAK52977	Human polynucleoti
37	103.5	9.5	4809	AAK51993	Human polynucleoti
38	101.5	9.3	2853	ABL29393	Drosophila melanog
39	101.5	9.3	6906	ABL29392	Drosophila melanog
40	101	9.3	914	AAK24074	Human EST R2810 DN
41	100.5	9.3	2826	ABL05376	Drosophila melanog
42	100	9.2	2564	ABL06127	Drosophila melanog
43	100	9.2	5340	ABL06126	Drosophila melanog
44	100	9.2	12403	ABL07646	Drosophila melanog
45	99.5	9.2	5643	ABK93133	Human prostate spe

ALIGNMENTS

RESULT 1
ID AA594693 standard; cDNA, 1340 BP.
AA594693;
12-MAR-2002 (first entry)
Rat secreted factor DNA clone P0184_D11 #1.
Rat; secreted factor polypeptide; cardiac disease; kidney;
inflammatory disease; congestive heart failure; myocarditis; ashma; ss;
dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
renal infection; hereditary nephritis; polycystic kidney disease;
chronic renal failure; renal vein thrombosis; medullary sponge kidney;
rheumatoid arthritis; osteoarthritis; poriasis; restenosis; PCR primer;

KW graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
 KW Alzheimer's disease; gene therapy.
 OS Rattus norvegicus.
 XX WO200174901-A2.
 PN 11-OCT-2001.
 XX 23-MAR-2001; 2001WO-US09555.
 PF 31-MAR-2000; 2000US-193548P.
 PR 14-MAR-2001; 2001US-0809545.
 XX (SCIO-) SCIOS INC.
 PA Stanton LW, White RT;
 XX WPI; 2002-010779/01.
 DR P-PSDB; AAU70146.
 XX Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX
 XX Claim 1; Fig 1; 189pp; English.
 PS
 CC The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 CC
 XX
 XX Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;
 SO
 Alignment Scores:
 Pred. No.: 9 436-89 Length: 1340
 Score: 1086.00 Matches: 203
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0
 US-09-809-545a-2 (1-203) x AAS94693 (1-1340)
 QY 1 MetThrAsnLysAlaValAsnProTyrThrAsnGlyTrrPylsLeuAsnProVal 20
 DB 535 ATGACTAATATAAAGCGCGTGAACCCCTACCAATGGCGGAAATTAATCAAGTTG 594
 QY 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaGln 40
 DB 595 GCGCGGCTTACAGCCCACTTCTATGACGACGCGTGTGTGGCCAGGCAACGAG 654
 QY 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
 DB 655 GAGGAGTCTTCATGATGACGTGGCCCAAGTTCTTGATATATCTTCTGCAATGCTTGGC 714
 QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
 DB 715 TTTCATATCCGCGCGCACTGCTGACGCTGATACGAGGGGCTCACTTGCAGGCGGT 774
 QY 81 GlyAlaThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100

DB 775 GGTCCGACCGTGTACACACCTTCAGAGCTGGGCGCCCCCAATCCCGGCTTAT 834
 QY 101 GlyGlyValValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
 DB 835 GCGGAGTGTGTATCAAGCCAGTGTATGGCAATTAATGTTACAGGTGTTACGCT 894
 QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
 DB 895 GCATACCGCTACGCGCCAGCCCACTCTGCTGCTGCTGCTACAGTACAGTTACGGA 954
 QY 141 ArgValTyrAlaAlaAspProTyrHisThrLeuAlaProAlaProThrTyrGlyVal 160
 DB 955 CGAGTTATGTCGCGACCCCTACACACACCTTGTCTTCAGCCCCCACTTACGCGCTT 1014
 QY 161 GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp 180
 DB 1015 GTGGCATTAATGCTTTTGGCGCTTGACCGATGCCAAGACAGGATGATGAT 1074
 QY 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200
 DB 1075 GTGGGCTGCTGTTCTTCTTCATTTGACGCTAGTATATACCAAGGGGATACAAACCGTTT 1134
 QY 201 AlaProTyr 203
 DB 1135 GCTCCATAT 1143
 RESULT 2
 AAH13824
 ID AAH13824 standard; cDNA; 1513 BP.
 AC AAH13824;
 XX 26-JUN-2001 (first entry)
 DT
 XX Human cDNA sequence SEQ ID NO:10786.
 DE
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 8; SEQ ID 10786; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide and an oligonucleotide comprising a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
XX
SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;

Alignment Scores:

Pred. No.:	8.6e-82	Length:	1513
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	92.96%	Indels:	1
DB:	22	Gaps:	1

US-09-809-545A-2 (1-203) x AAH13824 (1-1513)

QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTPDLSLeuAsnProValVal 20

DB 806 ATGCAATAAATAAGAACCGCTACACCTTATCAATATGCTGMAATTGATTCAGTTGTG 865

QY 21 GAlaValAlaTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGAlaAlaAsnGln 40

DB 866 GGTGACGCTTACAGTCCCAATTTCTATGACGACGCTGCTGTTGCGACAGCCACACAG 925

QY 41 GAluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMet-ProGly 60

DB 926 GAGGATCTTCATGATACAGTGCCTCCAGTTCATCTGTATATCTTGCATATGCCAGGC 985

QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrAlaGlyAlaHisLeuArgGlyArg 80

DB 986 TTCCCGTATCCAGACGCCCGCCGCGCTACCGAGGGGCCACCTCGAGGCCGC 1045

QY 81 GAlaTyrThrValTyrAsnThrPheArgAlaAlaAlaProProProGlyLeuProAlaTyr 100

DB 1046 GGTGACACCGGTACACACCTTCAGGCGCGGCGCCGCCGCCCAATCCGCGCTAC 1105

QY 101 GAlaGlyAlaValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAla 120

DB 1106 GCGCGTGTGTTTACCGAGATGATTTATGTGAGAC---ATTATGCTGTATATCT 1162

QY 121 AAlaTyrAlaGlyTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140

DB 1163 GCATATCCCTACGCCACCTTACCTCCACTGCGCTGCTACAGTACGATACCGA 1222

QY 141 ArgValTyrAlaAlaAspProTyrHisThrLeuAlaProAlaProThrTyrGlyVal 160

DB 1223 CGAGTTATGCTGCGCGACCCCTACACCGACTGCTCCAGCCGCCACCTACGCGCTT 1282

QY 161 GAlaAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180

DB 1283 GGTGCAATGATGCTTTGCACTTGTACTGATGCCAAGACATGAGACCATCTGATGAT 1342

QY 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200

DB 1343 GTGGGCTCGTCTTCTTCAATGACAGCTAGTATATACCGAGGGGATACACCGTTT 1402

QY 201 AAlaProTyr 203

DB 1403 GCTCCATAC 1411

RESULT 3

AAK52245
ID AAK52245 standard; cDNA; 1800 BP.

XX AAK52245;

AC 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 790.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX tissue growth factor; immunomodulatory; cancer; leukemia;

XX nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

XX NO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693325.

XX 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AD, Yang Y, Wejrtman T, Goodrich R;

XX WPI; 2001-476283/51.

XX P-PSDB; AAM79112.

XX Nucleic acids encoding polypeptides with cytokine-like activities,

XX useful in diagnosis and gene therapy -

XX Claim 1; Page 2643-2645; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK51435) and the

XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to

XX cytokine, cell proliferation or cell differentiation or which may induce

XX production of other cytokines in other cell populations. The

XX polynucleotides and polypeptides are useful in gene therapy, vaccines or

XX peptide therapy. The polypeptides have various cytokine-like activities,

XX e.g. stem cell growth factor activity, haematopoiesis regulating

XX activity, tissue growth factor activity, immunomodulatory activity and

XX treatment of cancer, leukemia, nervous system disorders, arthritis and

XX inflammation.

XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666

XX (AAM80020) are omitted as the relevant pages from the sequence listing

XX were missing at the time of publication.

XX

SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;

XX

Alignment Scores:

Pred. No.:	1.06e-81	Length:	1800
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	92.96%	Indels:	1
DB:	22	Gaps:	1

US-09-809-545A-2 (1-203) x AAK52245 (1-1800)

QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTPDLSLeuAsnProValVal 20

```

Db 1093 ATGACAAATAAAGACCCCTCAACCTTATACAAATGCTGGAAATGAATCAGTTGTG 1152
Qy 21 G1YAlaValTYrSerProAspPheTYrAlaG1YThValLeuLcYag1n1a1aAngin 40
Db 1153 GGTGCGATCTTACAGTCCCAATTTCTATGAGGACGCGTCTGTGTGCGGACCAACGAG 1212
Qy 41 GIUGlySerSerMetTYrSerGlyProSerSerLeuValTYrThSerAlaMetProGly 60
Db 1213 GAGGATCTTCCATGTACAGTGGCCCCCAGTTCACTGTATATATCTTCTGCAATGCCAGGC 1272
Qy 61 PheProTYrProAlaAlaThraAlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArg 80
Db 1273 TTCCCGATCCAGCAGCAGCCCGCGCGCGCTTACGAGGGCGCAGCCTGCGAGGCCGC 1332
Qy 81 GIYArgThValTYrAsnThrPheArgAlaAlaAlaProProProProLleProAlaTYr 100
Db 1333 GGTCCGACCGGTACACACCTTCAGGGCGCGGGCCCCCGCCGATCCCGGCTTAC 1392
Qy 101 GIYGIYValValTYrGlnGluProValTYrGlyAsnLYaLeuLeuGlnGlyTYrAla 120
Db 1393 GCGGCTGTCTTACAGAGATGATTTATGTGCGAGAC--ATTATGTGTGTTATGCT 1449
Qy 121 AlaTYrArgTYrAlaGlnProThrProAlaThraAlaAlaAlaTYrSerAspSerTYrGly 140
Db 1450 GCATACCGCTACGCCAGCCTACCCCTGCGCAGTCCGCTGCTACAGTACGGA 1509
Qy 141 ArgValTYrAlaAlaAspProTYrHisH1sThrLeuAlaProAlaProThrTYrGlyVal 160
Db 1510 CAGATTATGCTGCGCAGCCCTTACCCAGCAGCAGCTTGTCCAGCCCCCTTACGGCGTT 1569
Qy 161 GIYAlaMetAsnAlaPheAlaProLeuThraAspAlaLYeThraSerHisAlaAsp 180
Db 1570 GGTGCGATTAAGTCTTTCACCTTGTACTGATGCAAGCTAGAGCCATGCTGATGAT 1629
Qy 181 ValGIYLeuValLeuSerSerLeuGlnAlaSerIleTYrGlnGlyGlyTYrAsnArgPhe 200
Db 1630 GTGGGCTCTGTTCTTTCATGTGAGGCTAGTATATACCGAGGGGAGATACAAACGTTTT 1689
Qy 201 AlaProTYr 203
Db 1690 GCTCCATAC 1698

```

RESULT 4
AAA07075 standard; cDNA; 2372 BP.

```

XX AC AAA07075;
XX DT 03-JUL-2000 (first entry)
XX DE CDNA encoding human ataxin-2 binding protein (A2BP).
XX KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
XX KW neuronal plasticity; cellular degeneration signal transduction pathway;
XX KW selective RNA transport; spinocerebellar ataxia type-2;
XX KW hyperproliferative disorder; ss.
OS Homo sapiens.
XX Key Location/Qualifiers
XX CDS 987..1979
XX FT /*tag= a
XX FT /product= "Human A2BP"
XX FT
XX WO2000012710-A1.
XX PD 09-MAR-2000.
XX PF 01-SEP-1999; 99WO-US20156.
XX PR 01-SEP-1998; 98US-0145391.
XX

```

(CEDA-) CEDARS SINAI MEDICAL CENT.
PI Pulst SM, Shibata H;
XX WPI: 2000-237873/20.
DR P-98DB; AA81462.
XX
PT Nucleic acid encoding an ataxin-2 binding protein useful for
PT inhibiting the expression of active proteins from the SCA2 gene for the
PT treatment of spinocerebellar ataxia type-2 -
XX
PS Claim 6; Page 74-77; 82pp; English.

This sequence represents cDNA encoding human ataxin-2 binding protein (A2BP). Nucleotide sequences encoding human A2BP were originally isolated in an adult brain cDNA library using the yeast two hybrid method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown function that is encoded by the SCA2 gene located on chromosome 12. SCA2 has been linked to the autosomal dominant neurodegenerative disorder spinocerebellar ataxia type-2. Individuals afflicted with the disease exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2 containing a polyglutamine stretch of about 35-39 residues, whereas that of normal individuals contains approximately 22 contiguous glutamine residues. A2BP and ataxin-2 are components of a cellular degeneration signal transduction pathway. The pathogenic expanded form of ataxin-2 has a higher affinity for A2BP relative to normal ataxin-2; the presence of the expanded form is likely to promote degeneration. A2BP and ataxin have also been found to have a role in gene regulation. The binding of A2BP to ataxin-2 plays an important role in controlling gene expression via the targeting of transport of specific RNAs, selective RNA transport being mediated via the RNA binding domains of A2BP. A2BP is expressed very early in embryonic development. Both ataxin-2 and A2BP are able to bind CC RNA, and are essential components of the RNA localisation network that establishes cellular polarity in embryogenesis. In highly differentiated, polarised cells such as neurons, A2BP and ataxin-2 have a similar function and are required for neuronal plasticity. A2BP nucleic acids may be used for the recombinant production of A2BP proteins or fragments thereof according to standard methodologies. For example, an A2BP protein CC with an ataxin-2 or RNA binding capability but no signal transduction CC function can be used as a dominant negative inhibitor of the cellular degeneration signal transduction pathway. A2BP proteins with a signal transduction function can be used to treat hyperproliferative disorders (e.g., cancer) via stimulation of the cellular degeneration pathway.

Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;

Alignment Scores:

Pred. No.:	1,21e-76	Length:	2372
Score:	955.00	Matches:	189
Percent Similarity:	86.94%	Conservative:	9
Best Local Similarity:	85.14%	Mismatches:	21
Query Match:	87.94%	Indels:	2
DB:	21	Gaps:	2

US-09-809-545A-2 (1-203) x AAA07075 (1-2372)

```

Qy 1 MetThraAsnLYeAlaValAsnProTYrThraSngLYrTYrLYeLeuAsnProValVal 20
Db 1572 ATGACAAATAAAGACCCCTCAACCTTATACAAATGCTGAAATGATCAGTTGTG 1631
Qy 21 G1YAlaValTYrSerProAspPheTYrAlaG1YThValLeuLcYag1n1a1aAngin 40
Db 1633 GGTGCGATCTTACAGTCCCAATTTCTATGAGGACGCGTCTGTGTGCGGACCAACGAG 1691
Qy 41 GIUGlySerSerMetTYrSerGlyProSerSerLeuValTYrThSerAlaMetProGly 60
Db 1692 GAGGATCTTCCATGTACAGTGGCCCCCAGTTCACTGTATATATCTTCTGCAATGCCAGGC 1751
Qy 61 PheProTYrProAlaAlaThraAlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArg 80
Db 1752 TTCCCGATCCAGCAGCAGCCCGCGCGCGCTTACGAGGGCGCAGCCTGCGAGGCCGC 1811
Qy 81 GIYArgThValTYrAsnThrPheArgAlaAlaAlaProProProProLleProAlaTYr 100

```


AC AAS93633;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #29437.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
PN MO200175067-A2.
PD 11-OCT-2001.
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
P-PSDB; ABG29446.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 29437; 103bp; English.XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and a
CC food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. MAS64197-MAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;

Alignment Scores:

Pred. No.:	Length:
4,64e-46	481
Score: 606.50	Matches: 120
Percent Similarity: 90.51%	Conservative: 4
Best Local Similarity: 87.59%	Mismatches: 10
Query Match: 55.85%	Indels: 3
DB: 23	Gaps: 1

US-09-809-545A-2 (1-203) x AAS93633 (1-481)

69 AATAAAATYrArgGlYAlaHisLeuArGglYAAGlYArGTThValTYrAsnThrPhe 88
Db 1 GGCGGCCGCTTACCGAGGGGGCGACCCTGCAGAGCCCGGTCGCAACCTGACAACACTTC 60
89 ATGTAATAAAlaProRobroProIleProAlaTYrGlYAlaValTYrGlnGluPro 108

Db 61 AGGGCGGCGGCGCCCGCCCGATCCCGGCTACGGCGGTGTGTTTACCAGGATGA 120
 Qy 109 VALTYRGIYASrLYLeuLeuGlnGlyGlyTYRAlaAlaTYRArgTYRAlaGlnProThr 128
 Db 121 TTTTATGGTGACAGAC--ATTATATGTGGTTATVGTGCTATACCGCTACGCCCTAC 177
 Qy 129 ProAlaThrAlaAlaAlaTYRSerAspSerTYRGIYArgValTYRAlaAla--AspProT 148
 Db 178 CCTGCACCTGCCGCTGCTACAGTACAGTTACGACAGATTATGACATTTTCCCCCT 237
 Qy 148 YRHAHisThrLeuAlaProAlaProThrTYRGIYValGlyAlaMetAsnAlaPheAlaP 168
 Db 238 GTTCACCACTACAGTCTCCAGCCCGCCACCTACGGCGTGTGCTCCATGATGCTTTTGCA 297
 Qy 168 IoleuThrAspAlaAlaYsThrArgSerHisAlaAspAspValGlyLeuValLeuSerSerL 188
 Db 298 CTTTACTGATGCTCCAAAGACTAGAGCCATGCTGATGATGTGGGTCTCGTTCTTTCTTCA 357
 Qy 188 euGlnAlaSerLeuTYRGIYArgTYRAsnArgPheAlaProTYR 203
 Db 358 TGCAGGCTAGTATATACGAGGGGATACACCGTTTGTCTCCATAC 404
 RESULT 7
 ID ABK34530/c
 AC ABK34530 standard; cDNA; 539 BP.
 DT ABK34530;
 DE 08-MAY-2002 (first entry)
 XX Human CDNA for novel secreted protein, SEQ ID 299.
 XX Human; ss; gene; secreted protein; immune deficiency; viral infection;
 XX bacterial infection; fungal infection; autoimmune disorder; burn;
 XX rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
 XX diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
 XX Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
 XX coagulation disorder; haemophilia; inflammatory disorder; ulcer;
 XX tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
 XX lymphoid cell deficiency.
 OS Homo sapiens.
 PN WO200177290-A2.
 PD 18-OCT-2001.
 PF 29-MAR-2001; 2001MO-US10295.
 PR 06-APR-2000; 2000US-194941P.
 PA (GEMV) GENETICS INST INC.
 PI Mong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ,
 PI Gulukota K, Graham JR;
 DR WPI; 2002-179323/23.
 XX Six hundred and twenty five polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders
 PS -
 PS Claim 1; Page 153; 339p; English.
 CC The invention relates to 625 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins, their complements and sequences that hybridise to them.
 CC Also included are a vector comprising the polynucleotide, a host cell
 CC transformed with the vector, the proteins encoded by the
 CC polynucleotides, antibodies that bind to the proteins and identification
 CC of modulators of the proteins or the expression of the polynucleotide.

CC The polynucleotides can be used as probes for the identification
CC and isolation of full length cDNA and genomic DNA. The polynucleotides
CC and proteins can also be used as nutritional supplements. The protein
CC is useful in the treatment of various immune deficiencies and disorders
CC such as viral infections, bacterial infections, fungal infections,
CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
CC autoimmune thyroiditis and diabetes) and allergic reactions and
CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haemopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.

SQ Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;

Alignment Scores:	
Pred. No.:	7.26e-46
Score:	605.00
Percent Similarity:	82.67%
Best Local Similarity:	80.67%
Query Match:	55.71%
DB:	24
Length:	539
Matches:	121
Conservative:	3
Mismatches:	7
Indels:	20
Gaps:	2

US-09-809-545A-2 (1-203) X ABK34530 (1-539)

[illegible]

RESULT 8
AAS66245/c
ID AAS66245 standard; cDNA; 567 BP.

AC AAS66245;
XX 13-FEB-2002 (first entry)
DT
DE
XX DNA encoding novel human diagnostic protein #2049.
XX
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss
KW

OS Homo sapiens.
XX
XX
PN WO200175067-A2.
XX
XX 11-OCT-2001.
PD
PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PL Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG02058.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 2049; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy technique to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 567 Bp; 92 A; 151 C; 169 G; 155 T; 0 other;

Alignment Scores:	
Pred. No.:	1,756-39
Score:	534.50
Percent Similarity:	60.62%
Best Local Similarity:	56.48%
Query Match:	49,22%
OB:	23
Length:	567
Matches:	343
Conservative:	8
Mismatches:	31
Indels:	5
Gaps:	5

US-09-809-545A-2 (1-203) X AAS66245 (1-567)

Oy	1	MettTranLysIyAlaIaValAsnProThyTrhTrhLengIYTrpVLeuAsnProValaI	20
Db	501	ATGACCAACAAGAMACGGGGAACCCCTACACCAACGGCTGGAACTTAATCCATGGTC	44
Oy	21	GIyAlaValIYrSerProAspPheTyAlaGlyTrhValLeuLeuCyGlnAlaIngIn	40
Db	441	GGCGGAGCTACGGCCCTGGAATTCATGCAAGTACG-----	40
Oy	41	GIuIySerSerMetTYrSerGIyProSerSerLeuValTYrThSerAlaMetProGIy	60
Db	405	-----GGG	40
Oy	61	PheProTYrProAlaAlaIaThrAlaAlaAlaIaTyTrAGIyAlaHisLeuArgGIyTrg	80

```

Db 402 TTCCCTACCCACCCGACAGCCGCTTCCCTACCGGGCGACATCTTGGGGCCGG 343
Qy 81 GYATGTTThValTYrAsnThrPheAglAlaAlaProProProIleProAlaTYr 100
Db 342 GGGCGGGCGGTATATATATATTTGGGCTGGCCACCCCATCCGACTTAC 283
Qy 101 GYGLYValValTYrGlnGlnProValTYrGlyAsnLYsLeuGlnGlnLYrTYrAla 120
Db 282 GGAGGCGGTGTATATATATATATTTATGGTGTAG--ATTATAGAGGCTACGCA 226
Qy 121 AATYrArgTYrAlaGlnProThrProAlaThrAlaAlaTYrSerAspSerTYrGly 140
Db 225 GCGTACAGATAGCTACGCTCAGCCGCT--GGAGCGCGCGAGCTTACGCGACGTTACGGC 169
Qy 141 ArgValTYr---AlaAlaAspProTYrHisThrLeuAlaProAlaProThrTYrGly 159
Db 168 AAGTCTTACGCGAGCTGCCACCGGTACATCACACCATCGGGCCGCGAGCTTACAGC 109
Qy 160 ValGlyAlaMetAsnAlaPheAlaPro----- 168
Db 108 ATTGAACCATGTGAACCTTCCACCGCTTCTTCGAGCATGAAGGCAAAACAA 49
Qy 169 -----LeuThrAspAlaLYrThrArgSerHis 177
Db 48 AAAACAAAAAATATCACAACAAAAAACAACAAAAACAC 10

RESULT 9
AAI61034/c
ID AAI61034 standard; cDNA; 939 BP.
XX
AC AAI61034;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5023.
XX
KW Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-44253/47.
XX P-PsDB: AAM41878.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX

```

```

PS Claim 1; SEQ ID NO 5023; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM3642-AAM4221) with nocotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, cancer diagnosis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other:
XX
Alignment Scores:
Pred. No.: 3,22e-39 Length: 939
Score: 534.50 Matches: 109
Percent Similarity: 60.62% Conservative: 8
Best Local Similarity: 56.48% Mismatches: 31
Query Match: 49.22% Indels: 45
DB: Gaps: 22
XX
US-09-809-545A-2 (1-203) x AAI61034 (1-939)
Qy 1 MetThrAsnLYsLYsAlaValAsnProTYrThrAsnGlyTYrLYsLeuAsnProValAla 20
Db 501 ATGACCAACAAGAAAGACGGGAGACCCCTTACACCAACGCGCTGGAACTAAATCCATGTGTC 442
Qy 21 GYAlaValAlTYrSerProAspPheTYrAlaGlyThrValLeuLYsCysGlnAlaAsnGln 40
Db 441 GCGGAGTCTTACGGGCTCGAATTTCTATGACATGAGC----- 406
Qy 41 GYGLYSerSerMetTYrSerGlyProSerSerLeuValTYrThrSerAlaMetProGly 60
Db 405 -----GGG 403
Qy 61 PheProTYrProAlaAlaThrAlaAlaAlaTYrArgLYsAlaHisLeuArgLYsArg 80
Db 402 TTCCCTACCCACCCACCGGACAGCCGCTTCCCTACCGGGCGGACATCTTGGGGCCGG 343
Qy 81 GYATGTTThValTYrAsnThrPheAglAlaAlaProProProIleProAlaTYr 100
Db 342 GGGCGGGCGGTATATATATATTTGGGCTGGCCACCCCATCCGACTTAC 283
Qy 101 GYGLYValValTYrGlnGlnProValTYrGlyAsnLYsLeuGlnGlnLYrTYrAla 120
Db 282 GGAGGCGGTGTATATATATATATTTATGGTGTAG--ATTATAGAGGCTACGCA 226
Qy 121 AATYrArgTYrAlaGlnProThrProAlaThrAlaAlaTYrSerAspSerTYrGly 140
Db 225 GCGTACAGATAGCTACGCTCAGCCGCT--GGAGCGGGCGAGCTTACGCGACGTTACGGC 169
Qy 141 ArgValTYr---AlaAlaAspProTYrHisThrLeuAlaProAlaProThrTYrGly 159
Db 168 AAGTCTTACGCGAGCTGCCACCGGTACATCACACCATCGGGCCGCGAGCTTACAGC 109
Qy 160 ValGlyAlaMetAsnAlaPheAlaPro----- 168
Db 108 ATTGAACCATGTGAACCTTCCACCGCTTCTTCGAGCATGAAGGCAAAACAA 49
Qy 169 -----LeuThrAspAlaLYrThrArgSerHis 177
Db 48 AAAACAAAAAATATCACAACAAAAAACAACAAAAACAC 10

RESULT 10
AAI61035/c

```


PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Qa, Zhou P, Goodrich R, Dermanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PSDB; AAM40093.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 1452; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX
SQ Sequence 918 BP; 235 A; 300 C; 235 G; 148 T; 0 other;

Alignment Scores:
Pred. No.: 3,47e-39 Length: 918
Score: 534.00 Matches: 105
Percent Similarity: 67.68% Conservative: 6
Best Local Similarity: 64.02% Mismatches: 23
Query Match: 49.17% Indels: 30
DB: 22 Gaps: 4

US-09-809-545A-2 (1-203) x AA159249 (1-918)
QY 1 MetThrAsnLysLeuValaValaenProTyrThrAsnGlyTyrLysLeuAsnProValaVal 20
DB 438 ATGACCAACAAAGAGAGCGGGAAACCCCTACACCAACGCGCTGAAGCTAAATCCAGTGC 497
QY 21 G1YAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaenGln 40
DB 498 GGGCGAGTCTACGGCGCTGAATCTTATGCACTGACG----- 533
QY 41 G1uglySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
DB 534 -----GGG 536
QY 61 pheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHsLeuArgGlyArg 80
DB 537 TTCCTCTACCCCAACCAACCGGCGCTGCTTACCGGGGCGCACATCTTCGGGGCGG 596
QY 81 G1YArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProTyrAlaTyr 100
DB 597 GGGCGGCGCTGATATATACATTTCGGGCTGCCACACCCCACTCCGACTTAC 656
QY 101 G1Yg1yValValTyrG1ng1yProValTyrG1yAsnLysLeuLeuG1ng1yTyrAla 120
DB 657 GGAAGCGGTGCTGATACAGATGATTTTATGTCGTAG--ATTATGAGGCTACGCA 713
QY 121 AlATyTArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
DB 714 GCCTACAGATACCTCGCCGCT--GCAGCGGCGGCACTTACAGCGACGTCAGTACGGC 770

QY 141 ArgValTyr---AlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGly 159
DB 771 AGAGTCTACGACGCTGCGCCGACCGTACCATCACACATCGGGCGCGGCACTTACGCG 830
QY 160 ValG1yAlaMet 163
DB 831 ATTGAACCAATG 842

RESULT 12
AA159248
ID AA159248 standard; cDNA; 1011 BP.
XX
XX AA159248;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 1451.
DE
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM leukaemia; se.
XX
XX Homo sapiens.
OS
XX WO200153312-A1.
PN
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Qa, Zhou P, Goodrich R, Dermanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PSDB; AAM40092.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 1451; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

Seq	Sequence	1011 BP; 260 A; 314 C; 267 G; 170 T; 0 other;
XX	Alignment Scores:	
XX	Pred. No.:	3.9e-39
XX	Score:	534.00
XX	Percent Similarity:	67.68%
XX	Best local Similarity:	64.02%
XX	Query Match:	49.17%
XX		22 Gaps: 4
US	US-09-809-545A-2 (1-203) x AAI59248 (1-1011)	
QY	1 MetThrAenlySLySAlaValAAspProCyTrhArnglyTTpLySLeAAsnProValVal 20	
Db	531 ATGACCAACGAAGAAGACGGGAGACCCCTTACACCAACGGCTGGAAGCTTAATCCAGTGGTC 5900	
QY	21 GLyAlaValAlTYrSerProAspPheTYrAlaGlyTrhValLeuLeuCyGSlnAlaSnGln 40	
Db	551 GCGCGACGCTACGGGCGCTGAATTTCTATGCAGTGACG----- 626	
QY	41 GluGlySerSerMetTYrSerGlyProSerSerLeuValTYrTrhSerAlaMetProGly 60	
Db	627 -----GGG 629	
QY	61 PheProTYrProAlaAlaTrhAlaAlaAlaAlaTYrArnglyAlaHisLeuArnglyArG 80	
Db	630 TTCCCTTACCCACCCACCGGACAGCCGCTTCTTACCGGGGCGACATCTTGGGGCGG 689	
QY	81 GLYArGTrhValAlTYrAsnTrhPheArGAlaAlaAlaProProProPoiLeProAlaTYr 1000	
Db	650 GCGCGGCGCGTGTATTAATACATTTCGGGCTCGGCACCCCCACCCTTCCAGCTTAC 749	
QY	101 GLyGlyValAlaTYrGngLpProValTYrGlyAsnLysLeuLeuGngLysGlyTYrAla 120	
Db	750 GGAAGCGTGGTATACAGATGGAATTTATGTGTGTAG--ATTATGGAGGCTTACGA 806	
QY	121 AlaTYrArngTYrAlaGlnProTrhProAlaTrhAlaAlaAlaTYrSerAspSerTYrGly 140	
Db	807 GCTTACAGATACGCTCACCGCGCT--GCACGGGCGGACGCTTACAGGACAGTTACGGC 863	
QY	141 ArgValAlTYr---AlaAlaAspProCYrHisTrhLeuAlaProAlaProThTYrGly 155	
Db	864 AGAGCTTACGCGACGCGGACCCGTCATCATACACCATCGGCGCGCGACCTTACAGC 923	
QY	160 ValGlyAlaMet 163	
Db	924 ATTGAACCATG 935	
RESULT 13		
AAS90510	AAS90510 standard; cDNA; 1252 BP.	
AC	AAS90510;	
XX	13-FEB-2002 (first entry)	
XX	DNA encoding novel human diagnostic protein #26314.	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX	Homo sapiens.	
XX	WO200175067-A2.	
XX	11-OCT-2001.	
XX	30-MAR-2001; 2001WO-US08631.	
XX	31-MAR-2000; 2000US-0540217.	
XX	23-AUG-2000; 2000US-0649167.	

PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR
XX P-PSDB: ABG26323.
DR
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 1; SEQ ID No 26314; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. A6564197-A6594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX
SQ Sequence 1252 BP; 330 A; 338 C; 324 G; 260 T; 0 other;

Alignment Scores:
Pred. NO.: 5.98e-37 Length: 1252
Score: 511.00 Matches: 119
Percent Similarity: 61.47% Conservative: 15
Best Local Similarity: 54.59% Mismatches: 40
Query Match: 47.05% Indels: 46
DB: Gaps: 7
US-09-809-545A-2 (1-203) x AAS90510 (1-1252)

QY 1 MethrAsenlyslsAlaValaenProTyrThrAsnGlyTrrpLysleuAsnProValaVal 20
Db 563 ATGACCAATAAAGAGATGTCACACCATATGCAAAATGGTGGAAATTAAGCCCACTAGTT 622
QY 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValleuLysGln----- 37
Db 623 GGAGCTGTAATGGTCCGGAGTTAATGACGACATCCAG-CTTATCAGCAGATGTGTCC 681
QY 38 -----AlaasnGlnGlySerSerMetTyrSerGlyProSerSerLeuValTyrThr 55
Db 682 TAGCGATGATGACGACGAGTCCCTCATCAGGAAGAGGGGGATCAACACTTCATTCTT 741
QY 56 SerAlaMetProGlyPheProTyrPro--AlaAlaThrAlaAlaAlaTyrArgly 74
Db 742 TA-ATCATTCCTGGCTTCCCTTACCTACTGACGACACACGCGACGCCCTTTCAGAGA 800
QY 75 AlaHisLeuArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaPropro 94
Db 801 GCCCATTTGAGGGGACAGGGGCGACAGTATATGTCAGTCCGA---CGGGTACCTCA 857
QY 95 ProProIleProAlaTyrGlyGlyValAlaTyrGln-GluProValTyrGlyAsnLysle 114
Db 858 ACAGCCATCCCCGCTTATCCAGGTGTGGTTTACACAGGACGATTTTACGGTGTGAC-- 915
QY 114 uleuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr----- 131

QY 116 GlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr----- 131
DB 1082 AGGGGTGATATGACGCTTACAGATATGACAGCCGTCTACGCAACCGCACCCGCT 1141
QY 132 -----AlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro 147
DB 1142 GCTCAGCCGCTGACGCGCTTACAGTATGAGTATGAGGAGGTGTACAGCCGCTAC 1201
QY 148 TyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAla 167
DB 1202 TAC---CATGCCCTTGCCCTCGCCCTGAGCTATGAGTGTGGCGCTGTG----- 1246
QY 168 ProLeuThrAspAlaValSerThrArgSerHisAlaAspAspValGlyLeuValLeuSerSer 187
DB 1246 ----- 1246
QY 188 LeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPheAlaProTyr 203
DB 1247 -----GCGAGTTTATACCGAGGTGCTACAGCCGATTTGCCCTTAC 1288
RESULT 15
AAS93634
ID AAS93634 standard; cDNA; 2118 BP.
AC AAS93634;
XX
XX 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #29438.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX W0200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HXSE-) HXSEQ INC.
PA Dirmnac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG29447.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 29438; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;
Alignment Scores:
Pred. No.: 3,66e-27 Length: 2118
Score: 405.50 Matches: 105
Percent Similarity: 37.38% Conservative: 9
Best Local Similarity: 34.43% Mismatches: 10
Query Match: 37.34% Indels: 181
DB: 23 Gaps: 7
US-09-809-545a-2 (1-203) x AAS93634 (1-2118)
QY 53 ValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAlaAlaTyr 72
DB 568 GTTGAATCAACTCAGTCCAGCCAGCTTCCCGTATCCAGAGCCACCGCGCGCTTAC 627
QY 73 ArgGlyAlaHisLeuAlaArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAla 92
DB 628 CGAGGGGGCACCTGCGAGGCGCGGTGCGACCGGTGTACACACCTTCAGGGCGCGCG 687
QY 93 ProProProProIleProAlaTyrGlyGlyVal----- 103
DB 688 CCCCCGCCCGCATCCGAGCTTACGCGGTCTGTGAATCTGTAGACTTCCGGAATAT 747
QY 104 -----ValTyr--- 105
DB 748 TTGCTTTGTCTGATGATGACTTTAACAATATGCTGCTGTGTGATCCGTGTATACT 807
QY 106 -----GlnGlu----- 107
DB 808 GGACAGCAGAACTAGCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 867
QY 108 -----ProValTyrGly 111
DB 868 CACTAAAATGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 921
QY 112 AsnTyrLeuLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
DB 922 CCAAGGCTTGTGTGAAGTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 981
QY 132 AlaAlaAlaTyrSerAsp----- 137
DB 982 GCGCGTCTACAGTACAGACAGTGTCACTGATGATGATGATGATGATGATGATGATGATG 1041
QY 137 ----- 137
DB 1042 AAGGAAAGAAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1101
QY 137 ----- 137
DB 1102 GGTGCTTACCAAGTTTACCAATTAGAAAGCAAGTTCAAGGGGTTCAGAGGTTCACAC 1161
QY 137 ----- 137
DB 1162 GCGTTTCATTTTCTATCCCACTTTCACCTTACACATGCTGCGAGTACATTAAACT 1221
QY 137 ----- 137
DB 1222 TTAGACAGTGGGATGCAATGCGACAGCAAAATAGACACTGCCAGCTTCATGAGATT 1281
QY 138 -----SerTyrGlyArgValTyrAlaAlaAspProTyrHisHisThr 151
DB 1282 TCGTTCCAGGAGGTGAAGTTTACGAGAGTATATGCTGCGACCTTACACACACCA 1341
QY 152 LeuAlaProAlaProThrTyrGlyValGlyValaMetAsnAlaPheAlaProLeuThrAsp 171

```

Db 1342 CTGCTCCAGCCCCCACTACGGCGTTGGTGCATGAT----- 1380
QY 172 ALALYSTRARGSERHISALASPAPVALGILYLEUVALLIUSERSERLEUGINAIASER 191
Db 1381 -----AGACAGTAC-----AGTTCTTGTTGTGTATCC 1407
QY 192 ILETYRGINGLYGLY 196
Db 1408 GTTGCACACTGTGTGA 1422

```

Search completed: March 16, 2003, 02:38:55
 Job time : 232 secs

• • • •

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 16, 2003, 02:29:52 ; Search time 1461 Seconds

(without alignments)
2250.297 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MINKKAVNPYTNCKMLNPV.....VLSSLQASIVQGGYRPAFY 203

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xld
-Q=/cgr2_1/USPTO.spool/US09809545/rnac 07032003_153853_11467/app_query.fasta_1.391
-DB=EST -QMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LDEPXT=0
-UNITS=bites -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09809545 @CNC 1.1 2874 @rnac 07032003_153853_11467 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAT -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hlc: *
9: gb_estl: *
10: gb_estc: *
11: gb_hlc: *
12: gb_estc: *
13: gb_estc: *
14: gb_estc: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_int: *
20: em_gss_pln: *
21: em_gss_vit: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rnd: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	982.5	90.5	4039	11 BC026312	BC026312 Homo sapi
2	804.5	74.1	533	12 BF549922	BF549922 UI-R-E-D
3	763.5	70.3	723	12 BC306387	BC306387 fms8d11.x
4	712	65.6	495	9 A1656928	A1656928 ct48e08.x
5	698.5	64.3	774	10 AV729057	AV729057 AV729057
6	691.5	63.7	773	10 AV729198	AV729198 AV729198
7	683	62.9	391	9 A1186272	A1186272 qd20b08.x
8	677.5	62.4	521	13 BM183349	BM183349 fv10e03.x
9	658.5	60.6	477	10 AM197589	AM197589 km44g02.x
10	658.5	60.6	480	9 A1401040	A1401040 ch27a12.x
11	628.5	57.9	449	9 A1374891	A1374891 ch27a12.x
12	628.5	57.9	456	9 A1917298	A1917298 cs96g09.x
13	622.5	57.3	383	9 A1189407	A1189407 qd01h05.x
14	617	56.8	607	9 AUI45000	AUI45000 AUI45000
15	603	55.5	439	9 A1968176	A1968176 w14a06.x
16	603	55.5	439	10 AM589795	AM589795 hg22d04.x
17	603	55.5	439	10 BE501380	BE501380 7a41n09.x
18	576.5	53.1	459	9 A1244212	A1244212 q186b11.x
19	569	52.4	420	9 A1655094	A1655094 wb67a10.x
20	554.5	51.1	427	12 BF223478	BF223478 7g33e01.x
21	534.5	49.2	367	9 A1095813	A1095813 qb20g11.x
22	534	49.2	458	9 A1811011	A1811011 cu10c11.x
23	534	49.2	494	9 A1669300	A1669300 wb85b02.x
24	534	49.2	547	14 BM717439	BM717439 UI-E-EJO-
25	534	49.2	555	14 BM714144	BM714144 UI-E-EJO-
26	532.5	49.0	721	14 BM716595	BM716595 UI-E-DX1-
27	530	48.8	576	14 BM674802	BM674802 UI-E-EJO-
28	530	48.8	610	14 BM727789	BM727789 UI-E-EJO-
29	528.5	48.7	664	10 BB427515	BB427515 BB427515
30	525	48.3	1589	11 BC002124	BC002124 Mus muscu
31	523.5	48.2	556	10 AM015681	AM015681 UI-H-B10P
32	521	48.0	392	9 AA975235	AA975235 c936c08.b
33	521	48.0	346	9 A1674243	A1674243 wc45a12.x
34	517	47.6	594	12 BG380409	BG380409 UI-R-CTO-
35	516	47.5	679	14 BM727918	BM727918 UI-E-EJO-
36	514	47.3	468	9 A1421028	A1421028 cf04d06.x
37	513	47.2	459	9 A1668768	A1668768 wc14a04.x
38	508.5	46.8	680	11 AK005186	AK005186 Mus muscu
39	506.5	46.6	372	9 A1799929	A1799929 wc41a05.x
40	500	46.0	773	13 B1856648	B1856648 603385804
41	495	45.6	690	14 BQ448092	BQ448092 UI-H-EU1-
42	481	44.3	495	9 A1682528	A1682528 wc54e02.x
43	478.5	44.1	443	9 AA773715	AA773715 af81c05.x
44	465	42.8	660	10 AM157266	AM157266 au93g01.x
45	462	42.5	525	9 A1390973	A1390973 mb94d11.y

ALIGNMENTS

RESULT 1
LOCUS BC026312 4039 bp mRNA linear HTC 08-APR-2002
DEFINITION Homo sapiens, clone IMAGE:4815500, mRNA.
ACCESSION BC026312
VERSION BC026312.1 GI:20070932
KEYWORDS
SOURCE HTC.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 4039)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian

REMARK
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgs.nci.nih.gov>
Contract: MGC help desk

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiroki Tohyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdexaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAX Plates: 32 Row: 1 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922072
This clone has the following problem: frame shifted.

FEATURES

Source

1. 4039
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4815500"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"

BASE COUNT 1201 a 858 c 876 g 1104 t
ORIGIN

Alignment Scores:

Pred. No.: 3.09e-92 Length: 4039
Score: 982.50 Matches: 189
Percent Similarity: 95.07% Conservative: 4
Best Local Similarity: 93.10% Mismatches: 9
Query Match: 90.47% Indels: 2
DB: 11 Gaps: 1

US-09-809-545a-2 (1-203) x BC026312 (1-4039)

QY 1 MetThAsnLysLysAlaValaenProTyThAsnGlyTrpLysLeuAnpProVala 20
DB 823 ATGACAAATTAAGAAAGACCGTCAACCTTATACAAATGCTGGAAATTGAATCCAGTTTG 882

QY 21 GtAlaValaTySerProAspPheTyAlaGlyThValLeuLeuCyGlnAlaenGln 40
DB 883 GGTGCGAGTCTACAGTCCCGAATCTATGACGACGCGTCTGTGTGCGACGCCAACG 942

QY 41 GtGlySerSerMetTySerGlyProSerSerLeuValTyThSerAlaMetProGly 60
DB 943 GAGGGATCTTCCATGACAGTGCCTCCAGTTCCTGTATATCTCTCGAATGCCAGGC 1002

QY 61 PheProTyProAlaAlaThraAlaAlaAlaAlaTyArGtAlaAlaAlaAlaAla 80
DB 1003 TTCCCTATCCAGACCCAGCCGCCGCCCTTACCGAGGGCGGACCTTCCAGGCCGC 1062

QY 81 GtArGtThValTyArAnThrPheArGAlaAlaAlaProProProProAlaProAlaTy 100
DB 1063 GGTGCGACCGTGAACAACCTTCAAGGGCGGGCGGCCGCCGCCGATCCCGGCTAC 1122

QY 101 GtGlyValaValTyGtGtGtProValTyGtAlaLysLeuLeuGtGtGtGtGtAla 120
DB 1123 GGGGAGTGTGTTTACAGGATGATTTATGTGTCAGAC---ATTATGATGTTATGCT 1179

QY 121 AlaTyArGtTyAlaGtInpProThProAlaThAlaAlaAlaAlaAlaAlaAlaAla 140
DB 1180 GCATTACCGCTACCGCAGCTACCTGCGACCTGCGCGCTGCTCAAGTACGACATTACGA 1239

QY 141 ArgValaTyAlaAlaAlaAspProTyThiShiThLeuAlaProAlaProThTyGtVal 160
DB 1240 CCAGTTTATGCTGCCGACCCCTTACACACGACGATGCTGCCAGC---CCGACTTACGCGCTT 1298

QY 161 GtAlaMetAsnAlaPheAlaProLeuThAspAlaLysThArSerPheAlaAsp 180
DB 1299 GGTGCGATGAATGCTTTGACCTTGTACTGATGCGCAAGACTAGAGCCATCTATGAT 1358

QY 181 ValGtLeuValLeuSerSerLeuGlnAlaSerIleTyGtGtGtGtGtAlaAlaAla 200
DB 1359 GTGGGCTGCTGCTTCTTCTTCACTGACGCTACTATATACGAGGGGATACACCGTTT 1418

QY 201 AlaProTy 203
DB 1419 GCTCATAC 1427

RESULT 2

BF549922/c BF549922 533 bp mRNA linear EST 12-DEC-2000

LOCUS UI-R-E0-bv-c-04-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone

DEFINITION UI-R-E0-bv-c-04-0-UI 5', mRNA sequence.

ACCESSION BF549922
VERSION BF549922.1 GI:11659652

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 533)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

AUTHORS

TITLES

JOURNAL

MEDLINE

COMMENT

SOURCE

FEATURES

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mesoare@blue.weeg.uiowa.edu
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Clones will be available through Research Genetics (www.reagen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNL (info@image.llnl.gov). IMAGE ID= 1777166
Seq primer: M13 Forward.

Location/Qualifiers

1. 533

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

BASE COUNT 134 a 125 c 164 g 109 t 1 others

ORIGIN

Alignment Scores: 9.72e-75 Length: 533
Pred. No.: 804.50 Matches: 155
Score: 804.50 Conservative: 4
Percent Similarity: 94.08% Mismatches: 5
Best Local Similarity: 91.72% Indels: 5
Query Match: 74.08%

DB: 12 Gaps: 2

US-09-809-545a-2 (1-203) x BF549922 (1-533)

Qy 39 AsnGlnGluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAla--- 57
|||||
Db 532 AACGAGGAGATCTTCATGATACAGTGGCCCGCAGTCTGATATATACCTTGCAAT 473
58 -----MetProGlyPheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGly 74
|||||
Db 472 TCTGTTCCAGGCTGCTGCTTCATATCCGGCCCGCAGCTGTCAGCTGACGAGG 413
472 TCTGTTCCAGGCTGCTGCTTCATATCCGGCCCGCAGCTGTCAGCTGACGAGG 413
Qy 75 AlaHisLeuArgGlyArgGlyArgThrValTyrAnthrPheArgAlaAlaAlaProPro 94
|||||
Db 412 GCTCACCCTTCAGGCGCGGTGCGCAGCGTACAAACCTTCAGAGCTGCGCGCCCA 353
412 GCTCACCCTTCAGGCGCGGTGCGCAGCGTACAAACCTTCAGAGCTGCGCGCCCA 353
Qy 95 ProProLeuProAlaTyrGlyGlyValValTyrGlnGluProValTyrGlyAsnLeu 114
|||||
Db 352 CCCCACATCCCGGCTATGGGGGTGTGTTTACAGAGATGATTTATGTGTGACAGAC-- 296
352 CCCCACATCCCGGCTATGGGGGTGTGTTTACAGAGATGATTTATGTGTGACAGAC-- 296
Qy 115 LeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAla 134
|||||
Db 295 ATTATAGTGTGTTACGTCGATACCGATACCGCCAGCCACCTGCGCATGCTGCTGCC 236
295 ATTATAGTGTGTTACGTCGATACCGATACCGCCAGCCACCTGCGCATGCTGCTGCC 236
Qy 135 TyrSerAspSerTyrGlyArgValTyrAlaAlaAspProTyrHisSerThrLeuAlaPro 154
|||||
Db 235 TACAGTGACAGTTACGAGCAGATTATGCTGCCAGCCCTACCAACACACACTGCTCCA 176
235 TACAGTGACAGTTACGAGCAGATTATGCTGCCAGCCCTACCAACACACACTGCTCCA 176
Qy 155 AlaProThrTyrGlyValGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 174
|||||
Db 175 GCCCCACATCCCGGCTATGGGGGTGTGTTTACAGAGATGATTTATGTGTGACAGAC-- 296
175 GCCCCACATCCCGGCTATGGGGGTGTGTTTACAGAGATGATTTATGTGTGACAGAC-- 296
Qy 175 ArgSerHisAlaAspAspValGlyLeuValLeuSerSerLeuGlnAlaSerHisTyrGln 194
|||||
Db 115 AGGAGCCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56
115 AGGAGCCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56
Qy 195 GlyGlyTyrAsnArgPheAlaProTyr 203
|||||
Db 55 GGGGATACACCGTTTGTCTCCATAT 29
55 GGGGATACACCGTTTGTCTCCATAT 29
RESULT 3
BG306387/c 723 bp mRNA linear EST 13-FEB-2002
LOCUS f558d11.x1 Zebrafish adult retina cDNA Danio rerio cDNA clone
DEFINITION 4199493.3 similar to TR:090UM3 Q90UM3 D14P2.2; mRNA sequence.
ACCESSION BG306387
VERSION BG306387.1 GI:13103914
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 723)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Maria,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Page,D., Harvey,N., Schuk,R., Ritey,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE Washu Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
Library constructed by: Chandra Tucker and Gregory Niemi DNA
Sequencing by: Washington University Genome Sequencing Center
distribution: RessourcenzentrumPrimatendank, Berlin, Germany
(web address: www.rzpd.de)
Seq primer: T7 from Glibco

High quality sequence stop: 472.

FEATURES
source
1..723
/organism="Danio rerio"
/strain="wild-type"
/db_xref="taxon:7955"
/clone="4199493"
/sex="mixed"
/dev_stage="1-2 years"
/lab_host="E.Coli XL1-Blue MRF" (XL1-Blue MRF)"
/note="vector: Lambda ZAP II (pbluescript SK-); Site_1:
ECORI; Site_2: SalI; This Zebrafish library was
constructed by Dr. Susan E. Brockerhoff (email:
sbrock@u.washington.edu) RZPD library number: 760"

BASE COUNT 153 a 179 c 190 g 200 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 3,11e-70 Length: 723
Score: 763.50 Matches: 149
Percent Similarity: 77.94% Conservative: 10
Best Local Similarity: 73.04% Mismatches: 16
Query Match: 70.30% Indels: 29
DB: 12 Gaps: 3

US-09-809-545a-2 (1-203) x BG306387 (1-723)

Qy 1 MetThrAsnLeuValAlaValAsnProTyrThrAsnGlyTyrPheLeuAsnProValAl 20
|||||
Db 581 TTGCGAATATATAGACAGCGCAACCATATGACATGCTGGAAGTGAATCCAGTGTG 522
581 TTGCGAATATATAGACAGCGCAACCATATGACATGCTGGAAGTGAATCCAGTGTG 522
Qy 21 GlyValAlaTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaAsnGln 40
|||||
Db 521 GGTCAAGTCTACAGCCAGCAATTTCTAT----- 495
521 GGTCAAGTCTACAGCCAGCAATTTCTAT----- 495
Qy 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
|||||
Db 494 -----GCAGTCCAGGC 483
494 -----GCAGTCCAGGC 483
Qy 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 79
|||||
Db 482 TTCCCATACCCAGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423
482 TTCCCATACCCAGACGACG 423
Qy 80 ArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaAlaAlaAlaAlaAlaAla 99
|||||
Db 422 AGAGGCGGACCGTTTCAACACGTTTGGGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 363
422 AGAGGCGGACCGTTTCAACACGTTTGGGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 363
Qy 100 TyrGlyGlyValValTyrGlnGluProValTyrGlyAsnLeuLeuGlnGlyTyr 119
|||||
Db 362 TATGAGGTGTGTTTACAGAGAGGGTTTACGTCAGAT---ATTATGGTGTGTTAC 306
362 TATGAGGTGTGTTTACAGAGAGGGTTTACGTCAGAT---ATTATGGTGTGTTAC 306
Qy 120 AlaAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyr 139
|||||
Db 305 ACTCCCTACCATGCTACGCTGCTACGCTGCTACGCTGCTACGCTGCTACGCTGCTAC 246
305 ACTCCCTACCATGCTACGCTGCTACGCTGCTACGCTGCTACGCTGCTACGCTGCTAC 246
Qy 140 GlyArgValTyrAlaAlaAspProTyrHisSerThrLeuAlaProAlaProTyrGly 159
|||||
Db 245 GGAGGAGTTTATGCTGCGACCGCTTACACACGCTTGTCTTCCAGACGCGCATACAGC 186
245 GGAGGAGTTTATGCTGCGACCGCTTACACACGCTTGTCTTCCAGACGCGCATACAGC 186
Qy 160 ValGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaAlaAlaAlaAlaAla 179
|||||
Db 185 GTTGTCATGATGATGCTTTCGACCTTATGATGATGATGATGATGATGATGATGATGAT 126
185 GTTGTCATGATGATGCTTTCGACCTTATGATGATGATGATGATGATGATGATGATGAT 126
Qy 180 AspValGlyLeuValLeuSerSerLeuGlnAlaSerHisTyrGlnGlyTyrAsnArg 199
|||||
Db 125 GATGTGGGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 66
125 GATGTGGGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 66
Qy 200 PheAlaProTyr 203
|||||
Db 65 TTCCGCGCATAT 54
65 TTCCGCGCATAT 54
RESULT 4


```

Qy 101 GYGLYVALValTYrGInGluProValTYrGlyAsnLYsLeuLeuGInGlyGlyTYrAla 120
Db 123 GCGCGTGTGTTTACAGAGATGATTTATGTCAGAC---ATTATGCGTATGCT 179
Qy 121 AATyrArgTYrAlaGInProThrProAlaThrAlaAlaTYrSerAspSerTYrGly 140
Db 180 GCATACCGCTACGCCACCTACCCCTGCCACTGCCGTGCTACAGTACAGTACGGA 239
Qy 141 ArgValTYrAlaAlaAspProTYrHisHisThrLeuAlaProAlaProThrTYrGlyVal 160
Db 240 CGAGTTTATGCTGCGACCCCTACACACGACGACTTGTCTCCAGCCCACTACGCGCTT 299
Qy 161 GYAlaMetAsnAlaPheAlaProLeuThraPalaLYsThraGSerHisAlaAsp 180
Db 300 GGTCCATGATGCTTTTGACACTTGATGCTCAAGACTAGAGCCATGCTGATGAT 359
Qy 181 ValGlyLeuValLeuSerSerLeuGInAlaSerIleTYrGInGlyLYsTYrAsnArgPhe 200
Db 360 GTGGGTCTCGTCTTTCTTCAATGACAGGCTATATACGAGGGGGATACACCGCTTT 419
Qy 201 AAlaProTYr 203
Db 420 GCTCCATAC 428

RESULT 6
LOCUS AV729198 773 bp mRNA linear EST 17-OCT-2000
DEFINITION AV729198 HTC Homo sapiens cDNA clone HTCBC09 5', mRNA sequence.
ACCESSION AV729198
VERSION AV729198.1 GI:10838619
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 773)
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
JOURNAL Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Source
1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBCD09"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 184 a 232 c 185 g 170 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1..3e-62 Length: 773
Score: 691.50 Matches: 131
Percent Similarity: 93.71% Conservative: 3
Best Local Similarity: 91.61% Mismatches: 8
Query Match: 63.67% Indels: 1
DB: 10 Gaps: 1
US-09-809-545a-2 (1-203) x AV729198 (1-773)

```

```

Qy 61 PheProTYrProAlaAlaThrAlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArg 80
Db 3 TTCCCGTATCCAGACGACCGTCGCGGCCGCTACGAGGGGGCGCACCTCGAGCGCG 62
Qy 81 GYArgThrValTYrAenThrPheArgAlaAlaAlaProPheProPheProAlaTYr 100
Db 63 GGTGCGACCGGTATCAACACCTTCAGGGCGCGCGGCCCGCCCGCCATCCGCGCTAC 122
Qy 101 GYGLYVALValTYrGInGluProValTYrGlyAsnLYsLeuLeuGInGlyGlyTYrAla 120
Db 123 GCGCGTGTGTTTACAGAGATGATTTATGTCAGAC---ATTATGCGTATGCT 179
Qy 121 AATyrArgTYrAlaGInProThrProAlaThrAlaAlaTYrSerAspSerTYrGly 140
Db 180 GCATACCGCTACGCCACCTACCCCTGCCACTGCCGTGCTACAGTACAGTACGGA 239
Qy 141 ArgValTYrAlaAlaAspProTYrHisHisThrLeuAlaProAlaProThrTYrGlyVal 160
Db 240 CGAGTTTATGCTGCGACCCCTACACACGACCTTGTCTCCAGCCCACTACGCGCTT 299
Qy 161 GYAlaMetAsnAlaPheAlaProLeuThraPalaLYsThraGSerHisAlaAsp 180
Db 300 GGTCCATGATGCTTTTGACACTTGATGCTCAAGACTAGAGCCATGCTGATGAT 359
Qy 181 ValGlyLeuValLeuSerSerLeuGInAlaSerIleTYrGInGlyLYsTYrAsnArgPhe 200
Db 360 GTGGGTCTCGTCTTTCTTCAATGACAGGCTATATACGAGGGGGATACACCGCTTT 419
Qy 201 AAlaProTYr 203
Db 420 GCTCCATAC 428

RESULT 7
LOCUS A1186273 391 bp mRNA linear EST 28-OCT-1998
DEFINITION q420b08.x1 Soares, placenta 8c09weeks 2NBHP8c09W Homo sapiens cDNA
ACCESSION A1186273
VERSION A1186273.1 GI:3736911
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 391)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1800 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 369.
FEATURES
Source
1..391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1724247"
/clone_lib="Soares placenta 8c09weeks 2NBHP8c09W"
/dev_stage="Two placenta: One from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pVT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGATGAGGAGCGCGCGCATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pVT73 vector

```

(Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 73 a 129 c 107 g 82 t

ORIGIN

Alignment Scores:

Pred. No.: 3.9e-62 Length: 391
Score: 683.00 Matches: 128
Percent Similarity: 99.23% Conservative: 1
Best Local Similarity: 98.46% Mismatches: 1
Query Match: 62.89% Indels: 0
Gaps: 0

US-09-809-545a-2 (1-203) x A1186273 (1-391)

QY 69 AAlaAlaAlaTyrArgGlyValAlaHisLeuArgGlyValArgThrValTyrAsnThrPhe 88
Db 1 GCGGCGCGCTACCGAGGGGCGCACTGCGAGGCGCGGTCGCACTGTACCAACACCTTC 60
QY 89 ArgAlaAlaAlaProProProProProProAlaTyrGlyValValTyrGlnGlnPro 108
Db 61 AGGCGCGGCG 120
QY 109 ValTyrGlyAsnValLeuLeuGlnGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThr 128
Db 121 GTGTATGGCAATAAATGCTGCGAGGTGTATGCTGCACTACCGCTACCGCGCTAC 180
QY 129 ProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyValTyrAlaAlaAlaProTyr 148
Db 181 CTGGCACTGCG 240
QY 149 HisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAlaPro 168
Db 241 CACCAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 169 LeuThrAspAlaAlaTyrThrArgSerHisAlaAspAlaValGlyLeuValLeuSerSerLeu 188
Db 301 TTGACTGAGTCCAGACACTGAGACCTGCTGAGTGTGGTCTCGTCTCTTCTTCATTG 360
QY 189 GlnAlaSerIleTyrGlnGlyGlyTyrAsn 198
Db 361 CAGGCTAGTATATACCGAGGGGATACAC 390

RESULT 8
LOCUS BM183349/c 521 bp mRNA linear EST 26-JUL-2002
DEFINITION fvl0e03.x2 zebrafish adult brain Dantio reio cDNA clone 5386085 3'
Similar to TR:09UGW3 09UGW3 D41P2.2 ; mRNA sequence.

ACCESSION BM183349
VERSION BM183349.1 GI:17514307
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Dantio reio

REFERENCE
AUTHORS Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kuebe, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittner, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE WASHU zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Other ESTs: fvl0e03.y1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenzentrumPrimatdatenbank, Berlin, Germany (web address: www.rzpd.de)
Seq primer: -40UP
High quality sequence stop: 395.
Location/Qualifiers
1..521
/organism="Dantio reio"
/db_xref="taxon:7955"
/clone="5386085"
/clone_id="zebrafish adult brain"
/sex="mixed male and female"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="E. coli DH10B"
/note="Vector: pZIRPLOX; Site_1: NotI; Site_2: SalI. Original library was constructed in lambdaZIRPLOX. Maas excision of the cDNA library was performed to yield pZIRPLOX plasmids. Insert check was done in original library."

FEATURES

source

BASE COUNT 111 a 124 c 149 g 137 t
ORIGIN
Alignment Scores:
Pred. No.: 2.22e-61 Length: 521
Score: 677.50 Matches: 135
Percent Similarity: 77.17% Conservative: 7
Best Local Similarity: 73.37% Mismatches: 13
Query Match: 62.38% Indels: 29
Gaps: 3

US-09-809-545a-2 (1-203) x BM183349 (1-521)

QY 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaAsnGln 40
Db 520 GGTGAGTCTACAGCCCGCAATTCAT 494
QY 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 493 -----GCAGTGCAGGCG 482
QY 61 PheProTyrProAlaAlaThr--AlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGly 79
Db 481 TTCCCATACCGACGACG 422
QY 80 ArgGlyArgThrValTyrAsnThrPheAlaAlaAlaAlaProProProProProAla 99
Db 421 AGAGCGCGGACCGCTACACACGTTTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 362
QY 100 TrrGlyGlyValValTyrGlnGlnProValTyrGlyAsnValLeuGlnGlnGlyTyr 119
Db 361 TTATGAGGTGTGTACCGAGCGGTTTACCGGTGACAGT--ATTATGATGTTAC 305
QY 120 AlaAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyr 139
Db 304 ACTGCTACCGATACACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
QY 140 GlyArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGly 159
Db 244 GGACGAGTTTATGCTGCGACCGCCCTACACACGCACTGCTGCGACCGCCACATACAC 185
QY 160 ValGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaAlaTyrArgSerHisAlaAsp 179
Db 184 GTTGTGCTCATGAATGCTTTCACCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125
QY 180 AspValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArg 199
Db 124 GATGTGGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 65

www.bio.1nl.gov/bdrr/image/image.html
Insert Length: 478 Std Error: 0.00
Seq primer: -40UP from Glibco.

FEATURES

source

1. 456
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:229168"
/clone_lib="NCI CGAP GC6"
/issue_type="pooled germ cell tumors"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and 8 circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 137 c 115 g 107 t
ORIGIN

Alignment Scores:

Pred. No.: 2.65e-56 Length: 456
Score: 628.50 Matches: 119
Percent Similarity: 94.57% Conservative: 3
Best Local Similarity: 92.25% Mismatches: 6
Query Match: 57.87% Indels: 1
DB: 9 Gaps: 1

US-09-809-545a-2 (1-203) x A1917298 (1-456)

Qy 75 AAlaHisleuArgGlyArgGlyValArgThrValTyrAenThrPheArgAlaAlaProPro 94
Db 1 GCGGACCTGCGAGCGCGGTCGCGACCGGTGACACACCTTCAGGCGCGCGCGCGCG 60
Qy 95 ProProIleProAlaTyrGlyGlyValValTyrGlnGlnProValTyrGlyAsnLysLeu 114
Db 61 CCCCCGATCCCGGCTACGCGCGGTGTTTACCGAGATGATTTATGATGTCGACAC--- 117
Qy 115 LeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAla 134
Db 118 ATTATGGTGGTATGTCGATACGCGACCGCCGACCCGCGACGCGCTGCC 177
Qy 135 TyrSerAspSerTyrGlyValValTyrAlaAlaAspProTyrHisHisThrLeuAlaPro 154
Db 178 TACAGTGCAGTTCAGCGAGCGAGTTATGCTGCCGACCCCTACACGACGACCTGCTCA 237
Qy 155 AlaProThrTyrGlyValAlaGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThr 174
Db 238 GCGCCGACCTACGCGGCTGGTGCCATGATGCTTTGACCTTTACGATGACCAAGACT 297
Qy 175 ArgSerHisAlaAspAspValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGln 194
Db 298 AGGAGCCATGCTGATGATGATGGGCTCTGCTCTTCTTCATTGACAGCTAGATATACCGA 357
Qy 195 GlyGlyTyrAsnAlaArgPheAlaProTyr 203
Db 358 GGGGGATACCAACCGTTTGTCTCCATAC 384

RESULT 13

LOCUS A1189407 383 bp mRNA linear EST 28-OCT-1998
DEFINITION g001h05.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1722489 3', mRNA sequence.

ACCESSION A1189407
VERSION A1189407.1 GI:3740616

KEYWORDS

EST.

SOURCE

human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 383)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.1nl.gov) for further information.
Insert Length: 486 Std Error: 0.00
Seq primer: -40UP from Glibco

High quality sequence stop: 381.
Location/Qualifiers

FEATURES

source

1. 383
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1722489"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"

/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'-
ACCTGAGAGATTCGCGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 69 a 122 c 102 g 90 t
ORIGIN

Alignment Scores:

Pred. No.: 8.88e-56 Length: 383
Score: 622.50 Matches: 118
Percent Similarity: 94.53% Conservative: 3
Best Local Similarity: 92.19% Mismatches: 6
Query Match: 57.32% Indels: 1
DB: 9 Gaps: 1

US-09-809-545a-2 (1-203) x A1189407 (1-383)

Qy 72 TyrArgGlyAlaHisleuArgGlyArgGlyValArgThrValTyrAenThrPheArgAlaAla 91
Db 1 TACGAGGGGGCGACCTGCGAGGCGCGGTGCGACCGGTGACACACCTTCAGGCGCGCG 60
Qy 92 AlaProProProIleProAlaTyrGlyGlyValValTyrGlnGlnProValTyrGly 111
Db 61 GCGCCGCGCGCGCGCGCGCTACGCGCGGTGTTTACCGAGATGATTTATGCT 120
Qy 112 AsnLysLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
Db 121 GCAGAC--ATTATGGTGGTATGTCGATACCGCTACGCGCGACCTTCGCGACT 177
Qy 132 AlaAlaAlaTyrSerAspSerTyrGlyValValTyrAlaAlaAspProTyrHisHisThr 151
Db 178 GCGCGCTGCTACAGTACAGTTCAGGAGGATTTATGCTGCCGACCCCTACACACCGCA 237
Qy 152 LeuAlaProAlaProThrTyrGlyValAlaGlyAlaMetAsnAlaPheAlaProLeuThrAsp 171
Db 238 CTGCTCAGCGCCGACCTACGCGCGGTGTCATGATGCTTTGACCTTTGACTGAT 297
Qy 172 AlaLysThrArgSerHisAlaAspAspValGlyLeuValLeuSerSerLeuGlnAlaSer 191
Db 298 GCCAAGACTAGACCAACGATGATGATGGGCTCTGCTCTTCTTCATTGACAGGCTAGT 357
Qy 192 IleTyrGlnGlyGlyTyrAsnArg 199
Db 358 ATATTCGAGGGGATACACCT 381

RESULT 14

A1145000/c

```

LOCUS       AUI45000                      607 bp    mRNA    linear    EST 05-AUG-2002
DEFINITION  AUI45000 HEMBA1 Homo sapiens cDNA clone HEMBA1003591 3', mRNA
ACCESSION   AUI45000
VERSION     AUI45000.1  GI:11006521
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 607)
AUTHORS     Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Salto,K., Kawai,Y.,
             Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
             S. and Isegai,T.).
TITLE       HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
             Salto,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
             Y., Nagai,T., Sugano,S., Isegai,T.)
JOURNAL     Unpublished (2000)
COMMENT     Contact: Takao Isegai
             Genomics Laboratory
             Helix Research Institute
             1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
             Tel: 81-438-52-3975
             Fax: 81-438-52-3986
             Email: genomics@hri.co.jp
             HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
             Research Institute; cDNA library construction; Department of
             Virology, Institute of Medical Science, University of Tokyo, and
             Helix Research Institute.
FEATURES    Location/Qualifiers
             source          1..607
                             /organism="Homo sapiens"
                             /db_xref="taxon:9606"
                             /clone="HEMBA1003591"
                             /clone_lib="HEMBA1"
                             /tissue_type="whole embryo, mainly head"
                             /dev_stage="embryo, 10 weeks"
                             /note="Vector: pME18SFL3"
BASE COUNT  140 a 152 c 179 g 129 t 7 others
ORIGIN
Alignment Scores:
Pred. No.:      6,47e-55      Length:      607
Score:          617.00      Matches:      136
Percent Similarity: 83.04%      Conservative: 6
Best Local Similarity: 79.53%      Mismatches: 21
Query Match:    56.81%      Indels:      8
DB:             9          Gaps:      2
US-09-809-545A-2 (1-203) x AUI45000 (1-607)
QY 40 GINGUGIYserSermetrYrserGly---ProSerSerleuValYrthrSerAla-Me 58
DB 600 CAGGAGGGATTTTTCATCGACAGNGCCCCCGAGTTCATGGAAANATTNTTGCAANT 541
QY 58 cProGlyPhePro-TyrProAlaAlaThraAlaAlaAlaAlaAlaAlaAlaAlaAla 78
DB 540 GCCAGGCTTCCGATTCAGAGAGCCACCGCGCGCGCTTCCAGAGGCGCCACTTGC 481
QY 78 TGGIYARG-GIYARGThrValYrAsnThrPheArgAlaAlaAla-ProProProProIi 97
DB 480 GAGNCGCGGGTGCACCGGTGACCAACACTTCAGGGCGCGCGCGCGCGCGCGAT 421
QY 97 ePro-AlaYrGly-GIYValValYrGInGluProValYrGlyAsnYsleuLeuGIn 116
DB 420 CCGCGGCTTACCGCGCGGTGTGTTCACAGATGATTTTATGTGCAGAC--ATTAT 364
QY 117 GIGIYTYrAlaAlaAlaYrArgYrAlaGInProThProAlaThraAlaAlaYrSer 136
DB 363 GGGGTTATGCTGCATACCGCTACGCCACCTTACCTGCTGCTGCTGCTGCTACGT 304
QY 137 AspSerTcGIYArgValYrAlaAlaAspProTYrHisThrleuAlaProAlaPro 156

```

```

DB 303 GACAGTTACGACGAGATTATGCTGCCAGACCCCTTACACACGACACTTGCTCCAGCCCC 244
QY 157 ThTYrGIYValGlyAlaMetAsnAlaPheAlaProleuThraSpAlaYrThrArgSer 176
DB 243 ACCTACGGCGTGTGGTCCATGATGATCTTTTGACACTTTTGACGATGCCAAGCTAGAGC 184
QY 177 HisAlaAspAspValGlyLeuValIeuSerSerLeuGInAlaSerIleTYrGInGlyGly 196
DB 183 CATGCTGATGATGTGGGTCTGCTTCTTCATGTGCGAGGTAGTATATACGAGGGGGA 124
QY 197 TYrAsnArgPheAlaProTYr 203
DB 123 TACAACCGTTTGCTTCATAC 103
RESULT 15
LOCUS       A1968176                      439 bp    mRNA    linear    EST 25-AUG-1999
DEFINITION  wu14806.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2516914 3',
             similar to TR:043251 O43251 HYPOTHETICAL 39.5 KD PROTEIN.; mRNA
             sequence.
ACCESSION   A1968176
VERSION     A1968176.1  GI:5764994
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 439)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cga@bcr-rt@mail.nih.gov
             Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
             R. Emmerit-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
             Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/ILMIL at:
             www.bio.liml.gov/bdtp/image/image.html
             Seq primer: -40UP from Glibco.
FEATURES    Location/Qualifiers
             source          1..439
                             /organism="Homo sapiens"
                             /db_xref="taxon:9606"
                             /clone="IMAGE:2516914"
                             /clone_lib="NCI CGAP GC6"
                             /tissue_type="Pooled germ cell tumors"
                             /lab_host="DH10B"
                             /note="Vector: pTRT3D-Pac (Pharmacia) with a modified
                             polylinker. Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                             from the normalized library NCI-CGAP GC4 was prepared, and
                             58 circles were made in vitro. Following HAP purification,
                             this DNA was used as tracer in a subtractive hybridization
                             reaction. The driver was PCR-amplified cDNAs from a pool
                             of 5,000 clones made from the same library (clonoids
                             1257096-1258631, 1469064-1470983, and 1475592-1476743).
                             Subtraction by Bento Soares and M. Fatima Bonaldo.
BASE COUNT  91 a 142 c 114 g 92 t
ORIGIN
Alignment Scores:
Pred. No.:      1,22e-53      Length:      439
Score:          603.00      Matches:      118
Percent Similarity: 88.15%      Conservative: 1
Best Local Similarity: 87.41%      Mismatches: 2
Query Match:    55.52%      Indels:      15
DB:             9          Gaps:      1
US-09-809-545A-2 (1-203) x A1968176 (1-439)

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 02:32:41 ; Search time 76 Seconds
(without alignments) 819.150 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTKKAVNPTNMGKLNPPV.....VLSSLOASLYQGGYNRPAPY 203

Scoring table: BL0SUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO.spool/US09809545/rnat 07032003.153853.11480/app_query.fasta_1.391
-DB=Issued_Patents_NA -QPM=fastacp -SUFFIX=p2n.rnt -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdl
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09809545 @CEN 1.61 @rnat 07032003.153853.11480 -NCPUE=6 -ICPU=3
-NO_XLXRY -NO_MMAT -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THEADS=1 -XGAPOP=10 -YGAPEXT=0.5 -Fgapop=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	955	87.9	2372	4	US-09-145-391-1
2	467.5	43.0	1506	3	US-09-176-657-5
3	101	9.3	914	2	US-08-935-450-10
4	98.5	9.1	821	4	US-08-990-823-62
5	96	8.8	3231	1	US-08-195-152-1
6	94	8.7	2368	4	US-09-343-011B-3
7	94	8.7	43280	2	US-08-804-227C-1
8	92.5	8.5	2371	2	US-08-343-443B-1
9	92.5	8.5	2412	1	US-08-437-027-18
10	92.5	8.5	28958	1	US-08-258-261B-6
11	92.5	8.5	28958	1	US-08-456-837-6
12	92.5	8.5	28958	1	US-08-457-342-6

13	92.5	8.5	28958	1	US-08-457-646A-6	Sequence 6, Appl1
14	92.5	8.5	28958	1	US-08-458-076A-6	Sequence 6, Appl1
15	92.5	8.5	28958	1	US-08-764-233A-4	Sequence 6, Appl1
16	92.5	8.5	28958	1	US-08-457-335A-6	Sequence 6, Appl1
17	92.5	8.5	28958	1	US-08-729-214-6	Sequence 6, Appl1
18	92.5	8.5	28958	3	US-09-028-934-6	Sequence 6, Appl1
19	92.5	8.5	49377	1	US-08-764-233A-1	Sequence 2, Appl1
20	92	8.5	53526	3	US-08-658-136-2	Sequence 2, Appl1
21	92	8.5	53577	3	US-08-658-136-1	Sequence 1, Appl1
22	91	8.4	1896	4	US-09-443-011B-4	Sequence 4, Appl1
23	90.5	8.3	10095	3	US-08-922-586-45	Sequence 4, Appl1
24	89	8.2	1140	3	US-09-023-173-4	Sequence 4, Appl1
25	88	8.1	2923	6	5187076-5	Patent No. 5187076
26	87.5	8.1	3155	4	US-09-442-100-7	Sequence 7, Appl1
27	87.5	8.1	44377	2	US-08-804-227C-7	Sequence 7, Appl1
28	87.5	8.1	44377	2	US-08-804-198-1	Sequence 1, Appl1
29	87	8.0	1352	1	US-08-552-142A-10	Sequence 10, Appl1
30	87	8.0	1452	1	US-08-552-142A-16	Sequence 16, Appl1
31	87	8.0	2418	1	US-08-462-184-1	Sequence 1, Appl1
32	87	8.0	2418	1	US-08-441-147-1	Sequence 1, Appl1
33	87	8.0	2418	5	PCT-US95-07536-1	Sequence 1, Appl1
34	86.5	8.0	2040	1	US-08-031-538-10	Sequence 10, Appl1
35	86.5	8.0	2888	4	US-08-765-907A-1	Sequence 1, Appl1
36	86.5	8.0	13807	4	US-09-052-469-5	Sequence 5, Appl1
37	86.5	8.0	14060	3	US-08-658-136-4	Sequence 4, Appl1
38	86.5	8.0	14148	4	US-09-052-469-7	Sequence 7, Appl1
39	86	7.9	735	4	US-09-336-536-47	Sequence 4, Appl1
40	86	7.9	1801	4	US-09-336-536-46	Sequence 46, Appl1
41	85	7.8	654	4	US-08-998-416-1144	Sequence 1144, Ap
42	85	7.8	1494	4	US-09-255-502-1	Sequence 1, Appl1
43	85	7.8	1497	1	US-08-322-677A-6	Sequence 6, Appl1
44	85	7.8	1497	1	US-08-322-676-6	Sequence 6, Appl1
45	85	7.8	1497	3	US-08-898-218-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-09-145-391-1
: Sequence 1, Application US/09145391
: Patent No. 6194171
: GENERAL INFORMATION:
: APPLICANT: Pulst, Stefan M.
: APPLICANT: Shibata, Hiroki
: TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
: TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
: FILE REFERENCE: CE 3093
: CURRENT APPLICATION NUMBER: US/09/145,391
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 2372
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (987)..(1979)
US-09-145-391-1

Alignment Scores:

Pred. No.: 1.3e-85 Length: 2372
Score: 955.00 Matches: 189
Percent Similarity: 86.94% Conservative: 4
Best Local Similarity: 85.14% Mismatches: 9
Query Match: 87.94% Indels: 21
DB: Gaps: 2

US-09-809-545A-2 (1-203) x US-09-145-391-1 (1-2372)

Qy 1 MetThrAsnLysLysGlnAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal 20
Db 1572 ATGCAATATAAAGACCGCTTATCAATGAGTGGAAATGCAATCCAGTTGTG 1631

Qy	21	GLYALAVALTYRSESRPROASPPHEPTYALIGLYTHNVALLEULEUCYSGINLAASNGIN	40
Db	1632	GGTGCAGCTCTKACAGTCCCGAATTTCTANCGACGACGGCTCTGTGTGCGACGGCCAAACNG	1691
Qy	41	GLUGLYSERSESRMCTYRSEGGYPROSESRLEUVALTYRTHSRNALAEWPROGGY	60
Db	1692	GAGGAGATCTTCACATGATACAGTGGCCCCCAGTTCACTTGATATATACCTTCGATATGCCAGGC	1751
Qy	61	PHETPROTYRPROALAAATHTHRAALAAALAAATYRARGGLVALAHISLEUKRGGLTYARG	80
Db	1752	TTCCGGATCCAGGAGGACCGCCGCCGCCCTACCGAGGGCGGACCTGGCAGAGCGCG	1811
Qy	81	GLYARGTHRVALTYRASNTHRPHEARGALAAALAPROPROBROPROIIEPROALATYR	100
Db	1812	GGTGGCACCGGTGACAAACACTTCAGGGCGGGCGGCCGCCGCCCGCATCCGGGCTAC	1871
Qy	101	GLYGLYVALVALTYRGLINGLUPROVALTYRGLYANLVLLEULEUNGGLYGLYTYRAL	120
Db	1872	GGCGGTGTGGTTTATCCAGGATGATTTATGTGTGCAGAC---ATTATGAGGTGTTATGC	1928
Qy	120	AAATYRARGYRVALAGINPROTHRPROALATHRAALAAALTYRSESRASPSER----	138
Db	1929	TGCATACCGCTACCGCCAGCTACCCCTCGACGTGGCGGCTCTACAGTGAACG-AAAATC	1987
Qy	139	-----TYRGLYARGVA	142
Db	1988	AGTTGCTTCTGCTGGCAGCAGATGAATTTCTTGTAACACCTCTGCAGTTATGAGCAGAT	2047
Qy	142	LYTRALAAALASPPROTYRTHSHIETHRLEUALAPROALAPROTHRTYRGLYVALGLYAL	162
Db	2048	TTATAGCTGCCACCCCTACACACCGACACTGTGCTCCAGGCCCCCACTACGGGGTGGTGC	2107
Qy	162	AMERASNALAPHEALAPROLEUTHRAAPALALVATHRARSGSERHIALAASPPASPVAGL	182
Db	2108	CATGAATGCTTTTCACCTTGACTGATGTCAGCAACCTAGAGCCATGCTGATGAATGTGG	2167
Qy	182	YLEUVALLEUSERLEUGINALASERILETYRGLINGLYGLYTYRASNARGPHEALAPR	202
Db	2168	TCTGCTCTTTCTTCATTTGACGGCTATATATACGAGGGGGATACAAACGTTTGGCTCC	2222
Qy	202	OTYR 203	
Db	2228	ATAC 2231	

RESULT 2
US-09-176-657-5

```

: Sequence 5, Application US/09176657
: Patent No. 6020164
:
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Tang, Y. Tom
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Lu, Aina
: APPLICANT: Baughn, Mariah R.
: TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
: FILE REFERENCE: PF-0611 US
: CURRENT APPLICATION NUMBER: US/09/176,657
: CURRENT FILING DATE: 1998-10-21
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PERL Program
:
: SEQ ID NO 5
:
: LENGTH: 1506
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: -
: OTHER INFORMATION: 1250374
:
: US-09-176-657-5

```

Alignment Scores:	1.8e-37	Length:	1506
Pred. No.:	467.50	Matches:	107
Score:			

```

Percent Similarity: 56.48% Conservative: 15
Best Local Similarity: 49.54% Mismatches: 39
Query Match: 43.05% Indels: 56
DB: 3 Gaps: 8

```

US-09-809-545A-2 (1-203) x US-09-176-657-5 (1-1506)

Oy	1	MethrksnlysvslalaValaIasnProYrtrthraAsglTrpIlyLeuAsnProValaI	20
Db	768	ATGACCAATAAAGAGATGGTCACACCAATTCGAAATGGTTGAAATTAAAGCCCACTT	827
Oy	21	GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaIasn---	39
Db	828	GGAGCTTATATATGTCGGAGATTATATGACATCCAGCTTCAAGCAGATGTGCTTA	887
Oy	40	---GlnGlnGlySerSerMetTyrSerGlyProSerSerIu-----ValTyrThrSer	56
Db	888	GGCAATATATCAGACAGTCCCTCATCGAAGAGGGGGATATACACTTACATTCCTTTA	947
Oy	57	AlaMetProGlyIlePheProTyrPro---AlaAlaThrAlaAlaAlaTyrAlaGlyAla	75
Db	948	ATCATCTCTGCTTCCTTACCTTACTGACGACACACGCGCGCTTTCAGAGAGACC	1007
Oy	76	HisLeuArgIlyArgIlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProPro	95
Db	1008	CATTGAGGGGGCAGAGGGCGGAGATATGTGGCAGTCCG---GGCGTACCTCAGCA	1064
Oy	96	ProIleProAlaTyrGlyGlyValValTyrGlnGluProValTyrGlyAsnIlyLeuLeu	115
Db	1065	GCCATCCCCCGCCATCC-----	1081
Oy	116	GlnGlnGlyIlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr-----	131
Db	1082	AGGGGTGATATACAGCTTACATATATGACAGCCGCTACTGCAACCGCAGCCAGCT	1141
Oy	132	-----AlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro	147
Db	1142	GCTGACACCGCTGACGCGCTTACATGACGGTATGCGAGGGTGTACACAGCCGACCC	1201
Oy	148	TyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAla	167
Db	1202	TAC---CATGCCCTTGCCCTCGCCGCTCACTATGAGATTGGCGGTGTG-----	1246
Oy	168	ProLeuThrAspAlaIlyThrTrArgSerHisAlaAspAspValGlyLeuValLeuSerSer	187
Db	1246	-----	1246

RESULT 3

```

US-08-935-450-10
? Sequence 10, Application US/08935450
? Patent No. 5977311
? GENERAL INFORMATION:
? APPLICANT: Nandabalan, Krishnan
? APPLICANT: Yang, Meijia
? APPLICANT: Schulz, Vincent
? TITLE OF INVENTION: 53P2 COMPLEXES
? FILE REFERENCE: 7934-054
? CURRENT APPLICATION NUMBER:
? CURRENT FILING DATE: 1997-09-23
? NUMBER OF SEQ ID NOS: 15
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 10
? LENGTH: 914
? TYPE: DNA
? ORGANISM: Homo sapiens
US-08-935-450-10

```

Alignment Scores:	
Pred. No.:	0.227
Length:	914

Score: 101.00 Matches: 48
Percent Similarity: 37.06% Conservative: 15
Best Local Similarity: 28.24% Mismatches: 59
Query Match: 9.30% Indels: 48
DB: 2 Gaps: 7

US-09-809-545A-2 (1-203) x US-08-935-450-10 (1-914)

```
Oy 24 TySerProaPheTyraGlyThrValLeuLeuCyGlnAlaAenGlnIuGlySer 43
    |||
Db 230 TACAGCCCT-----GCTCGAACCCTCCAGGGGCC 259
Oy 44 SerMetTySerGlyProSerSerLeuValTyThrSerAla-MetProGlyPheProTy 63
    |||
Db 260 AGCAGCTACATAGAACAGACGACATCCCTGCTCAAGCCCAATACAGACACCCACC 319
Oy 63 rProAlaAlaThrAla-----AlaAlaAlaTyraGlyAlaHisLeuAlaGlyAlaGly 81
    |||
Db 320 GTCAGCAGCTACAGCCCTTCCACAGCCGAGTTACAGCCAGCCACCTTAC--AACAGGG 376
Oy 81 YArgThrValTyraSerThrPheArgAlaAlaAla-ProProProPheProAlaTyrg 101
    |||
Db 377 GAGGTACAGCCAGGAGTTACAGAGCCACCGCTTCCACCTCCACACCACTGCTTACA 436
Oy 101 YGlyValValTyrgInGluProValTyrgIYasnIYsLeuLeuGlnIYgIYTyraAla 121
    |||
Db 437 AC-----TATGGAGC-----TACGGCGTTACACC 463
Oy 121 IaTyraGlyAlaGlnProThrProAlaThrAlaAlaAlaTySerAspSerTyrgIY 141
    |||
Db 464 CGGCCCCCTATACCCACAGCCGCCACCCGCCACGACAGACCTAC----- 507
Oy 141 rGValTyraAlaAlaAspProTyrHisIsthLeuAlaProAlaProThrTyrgIYValG 161
    |||
Db 508 -----CTCAGCCCAACTATACCACT 529
Oy 161 YAlaMetAsnAlaPheAlaProLeuThrAspAlaIYsThrArgSerHisAlaAspAlv 181
    |||
Db 530 ATCAGCACTATAGCCAGCAGTGGAACAGTACTACATAGAACAGCGGCCAGT----- 579
Oy 181 aIGlyLeuValLeuSerSerLeuGln 189
    |||
Db 580 --GGCGCATCTACTACGGAGACTACGA 603
```

RESULT 4

US-08-990-823-62
; Sequence 62, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990, 823D
; CURRENT FILING DATE: 1997-12-15
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254
; EARLIER FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 62
; LENGTH: 821
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: Modified base
; OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-62

Alignment Scores:

Pred. No.: 0.348 Length: 821
Score: 98.50 Matches: 59

Percent Similarity: 28.46% Conservative: 15
Best Local Similarity: 22.69% Mismatches: 95
Query Match: 9.07% Indels: 91
DB: 4 Gaps: 9

US-09-809-545A-2 (1-203) x US-08-990-823-62 (1-821)

```
Oy 1 MetThrAsnIYsYsAlaValaAsnProTyThrAsnGlyTrpIYsLeuAsnProVala 20
    |||
Db 70 ATGGCCAGCAGCGTGTGCGCCAGCCGACCCCAATCCGGCATGTGCAACCGGTCCG 129
Oy 21 GYAlaValaIYserProaPheTyraGlyThrValLeuLeuCyGlnAlaAenGln 40
    |||
Db 130 GGCAGCGCGTTCGAGCCGAGCGAGCGCTGGCGGTATCAGTCCGTGCGCTTCAACCCG 189
Oy 41 GluIYserSerMetTyThrSerGlyProSer-----SerLeuValTy 54
    |||
Db 190 AGGCGTGGGAGGAGACTCAAGTCCAGCCGCTGTCTCCAAACCCGGCTGTCTAC 249
Oy 55 ThrSerAlaMetProGlyPheProTyProAlaAlaThrAlaAlaAlaTyraGly 74
    |||
Db 250 TCGATGTGTGCGCCCAACCAACCGCCGCTACACGACTCCGCGGCGACCGCGGA 309
Oy 74 ----- 74
Db 310 GGAAGGCGCCGNTCGGATNCAACGCGTTGCNACGCGCGCTGCTTCNGATTGA 369
Oy 75 -----AlaHisLeuArgIYArg-----GlyArgThrValTy 85
    |||
Db 370 CCGGCGACGTAACCCCGGTATGGCAGCTTACGGAGAACAACTGGCCGCGCACGCGCAC 429
Oy 86 AsnThrPheArgAlaAlaAlaProProProPheProAlaTyrgIYValaIY 105
    |||
Db 430 CTCGCGCTGTACAGTTACCGCCCGAGCCGAGCCGCGCGCTGTGT----- 480
Oy 106 GlnGluProValTyrgIYAsnIYs-----LeuLeuGlnIYgIY 118
    |||
Db 481 -----GGTTCGCGCGCGCGCATCTGTCTTCAAGAGAGACGCGCATTCATCTA 534
Oy 119 -----TyraAlaAlaTyraGlyTrpIYAlaGlnIYProThrProAlaThr 131
    |||
Db 535 CGGCCAGTCCCTGAAATCTGACGTGGGGCGTACCGGCCGAGCCGCGCATTCACACT 594
Oy 132 AlaAlaAlaTySerAspSerTy----- 139
    |||
Db 595 GGGCGAGTATTTCCGATCGACATCGGACCGCAACCGCGTGGCCCATTCGCGTTCC 654
Oy 140 -----GlyArgValTyraAlaAspProTyrHis----- 150
    |||
Db 655 GCTGGCTGGCGCGCGCGGAGCGCGCATGTGCGCG-----CATTTGCGC 699
Oy 151 -----ThrLeuAlaProAlaProThrTyrgIY 159
    |||
Db 700 CTATGACCCGCACTGAGCCCTGAGCAATGTTGCGCTTCAACCCGCGGTTCCGCT 759
Oy 160 ValGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaIYsThrArgSerHisAlaAsp 179
    |||
Db 760 GCTGGAATCTCGACGCGGTGATCGGTACGACACCGGTGTGATGACATCGCAC 819
```

RESULT 5

US-08-195-152-1
; Sequence 1, Application US/08195152
; Patent No. 5679541
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Teest, Albritton & Herbert
; STREET: 4 Embardadero Center, Suite 3400

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2371 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 25..1992
US-08-343-443B-1

Alignment Scores:
Pred. No.: 5.75 Length: 2371
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 8.52% Indels: 41
DB: Gaps: 9

US-09-809-545A-2 (1-203) x US-08-343-443B-1 (1-2371)

QY 29 TTTAAGTCTThValLeuLeuCySGlnAla---AaNGlGlnGlySerSerMetTyrSer 47
DB 109 TATGCACAGACACC-----CAGGCATATGGCAACAAAGCTATGGAACCTATGGA 159
QY 48 GTPProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaIaThr 67
DB 160 CAGCCACGATGTCAGCTATACCCAGGCTCAGACACGCTGCAACCTATGGGACGAGCCG 219
QY 68 AAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyrAsnThr 87
DB 220 TATGCACCTCTTAT-----GGACAGCTCCCACTGCTTACTACT 261
QY 88 PheArgAlaAlaAlaProProProProIleProAlaTyrGlyValValTyrGlnGlu 107
DB 262 -----CCAACTGCCCCCCAG-----GCATACAGCCAG 288
QY 108 ProValTyrGlyAsnLeuLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnPro 127
DB 289 CTTGTCAGGGAT-----GGCAGCTGCTTATGATACCAACCACTGCT 333
QY 128 ThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro 147
DB 334 ACAGTCACCAACCCAGCC-----TCCATATGACCTCAGTCT 372
QY 148 TyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAla 167
DB 373 GCATATGGCACTCAGCTCTTATCCAGCTTATGGCAG-----CAG 414
QY 168 ProLeuThrAspAlaAlaTyrThrArgSerHisAlaAspAspValGlyLeuValLeuSerSer 187
DB 415 CCAAGCAGCAGCTCAGCTATCAAGACCGCAGAGTGAAGCAAGCCCACTGAGACTAATCAA 474
QY 188 LeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArg 199
DB 475 CCTCAATCTAGC-----ACAGGGGGTTACAAACGAG 504

RESULT 9
US-08-437-027-18
Sequence 18, Application US/08437027

Patent No. 5670317
GENERAL INFORMATION:
APPLICANT: Landanyi, Marc
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
TITLE OF INVENTION: SMALL ROUND CELL TUMOR
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/437,027
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 46416/JPW/CCA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-391-0525
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2412 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-437-027-18

Alignment Scores:
Pred. No.: 5.89 Length: 2412
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 8.52% Indels: 41
DB: Gaps: 9

US-09-809-545A-2 (1-203) x US-08-437-027-18 (1-2412)

QY 29 TTTAAGTCTThValLeuLeuCySGlnAla---AaNGlGlnGlySerSerMetTyrSer 47
DB 109 TATGCACAGACACC-----CAGGCATATGGCAACAAAGCTATGGAACCTATGGA 159
QY 48 GTPProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaIaThr 67
DB 160 CAGCCACGATGTCAGCTATACCCAGGCTCAGACACGCTGCAACCTATGGGACGAGCCG 219
QY 68 AAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyrAsnThr 87
DB 220 TATGCACCTCTTAT-----GGACAGCTCCCACTGCTTACTACT 261
QY 88 PheArgAlaAlaAlaProProProProIleProAlaTyrGlyValValTyrGlnGlu 107
DB 262 -----CCAACTGCCCCCCAG-----GCATACAGCCAG 288
QY 108 ProValTyrGlyAsnLeuLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnPro 127
DB 289 CTTGTCAGGGAT-----GGCAGCTGCTTATGATACCAACCACTGCT 333
QY 128 ThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro 147
DB 334 ACAGTCACCAACCCAGCC-----TCCATATGACCTCAGTCT 372

Qy 148 TyriShiEtThLeuAlaProAlaProThryTgIyValGlyAlaMeCAsnAlaPheAla 167
Db 373 GCATATGCACTCAGCCCTTATCCAGCTTATGGCAG-----CAG 414
Qy 168 ProLeuThraPAlaLalThraPArgSerHisAlaAspValGlyLeuValLeuSerSer 187
Db 415 CCAGCAGCCACTGCACTTACAAAGACCCGACGATGGAACCAAGCCCACTGAGACTAGTCAA 474
Qy 188 LeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArg 199
Db 475 CCTCATCTAC-----ACAGGGGGTTACCAACGAG 504
RESULT 10
US-08-258-261B-6
Sequence 6, Application US/08258261B
Patent No. 5639949
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258.261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6
Alignment Scores:
Pred. No.: 169 Length: 28958
Score: 92.50 Matches: 34
Percent Similarity: 41.28% Conservatave: 11
Best Local Similarity: 31.19% Mismatches: 52
Query Match: 8.52% Indels: 12
DB: 1 Gaps: 4

US-09-809-545a-2 (1-203) x US-08-258-261B-6 (1-28958)
Qy 47 SerGlyProSerSerLeuValTyrThSerAlaMeCProGlyPheProTyrProAlaAla 66
Db 10706 AGCGGCTCTCGACGATGGGATCGCTTCCGTCGAGGTACGCCCATCCGTCTCA 10765
Qy 67 ThrAlaAlaAlaAlaTyr---ArgGlyAlaHis-LeuArgGlyArgGlyArgThrValTy 85
Db 10766 CGCTCGCCCTCGGAGACCTCGACGCGCTCACCGCTGATCCGCTGCTCGGCTCCA 10825
Qy 85 rAsnThrPheArgAlaAlaAlaPProPProPProIleProAlaTyrGlyValValTy 105
Db 10826 TTCGACG---AGAAAGAGCCACCTCCCGCTGCTCTCTCCGCGGAGGCTCTTA 10882
Qy 105 rGlnGluProValTyrGlyAsnLysLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAl 125
Db 10883 CCCGAGGCT-----CGCGCTCGACGTGAAAGACTTCTTCCGCTACGCTCCCC 10933
Qy 125 aGln-----ProThrProAlaThrAlaAlaAlaTyrSerAspSerTy 139
Db 10934 GCAAGGTCTCCCTCCACCTACCTTCCAGCGAGAGCGGTCTTGCTGACGCTTCCA 10993
Qy 139 rGlyArgValTyrAlaAlaAspPro 147
Db 10994 CGGACGAACGCTTCCGACGTGCT 11018
RESULT 11
US-08-456-837-6
Sequence 6, Application US/08456837
Patent No. 5643774
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456.837
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:


```

: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,646A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CI13
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-457-646A-6

Alignment Scores:
Pred. No.: 169 Length: 28958
Score: 92.50 Matches: 34
Percent Similarity: 41.28% Conservative: 11
Best Local Similarity: 31.19% Mismatches: 52
Query Match: 8.52% Indels: 12
DB: 1 Gaps: 4

US-09-809-545A-2 (1-203) x US-08-457-646A-6 (1-28958)
Qy 47 SerGlyProSerSerLeuValTyrrThrSerAlaMetProGlyPheProTyrrProAlaAla 66
Db 10706 AGCGGCTCTCGAGCATGGGATCGCTTCTCGGTGAGGTGAGCCCATCCCGGTCTCA 10765
Qy 67 ThrAlaAlaAlaAlaTyrr--ArgGlyAlaHis-LeuArgGlyArgGlyArgThValTy 85
Db 10766 CGCTCGCGCTCGGAGCATGTCGAGCGCTCACCCCTCGATCCGTCGTCGCGCTCCA 10825
Qy 85 rAsnthrPheArGlyAlaAlaAlaProProProProleProAlaTyrrGlyGlyValTy 105
Db 10826 TTCGACG--AGAGAGAGCCACCTCGCCGCTCTCTCTCTCGGGCGGAGCTCTCA 10882
Qy 105 rGlnGluProValTyrrGlyAsnLysLeuLeuGlnGlyGlyTyrrAlaAlaTyrrArgTyrrAl 125
Db 10883 CCGGAGCGCT-----CGGCTCGACGTGAGAGACTTCTCGCGCCCTACGCTCCCC 10933
Qy 125 agIn-----ProThrProAlaThrAlaAlaAlaTyrrSerAspSerTy 139
Db 10934 GCAAGGTCTCCCTCCCACTACCCCTTCACGCGAGCGGTTCTGCTCGACGTCTCCA 10993
```

```

Qy 139 rGlyArgValTyrrAlaAlaAspPro 147
Db 10994 CGAGACGACGCTTCGACGCTCGCT 11018

RESULT 14
US-08-458-076A-6
: Sequence 6, Application US/08458076A
: Patent No. 5698425
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,076A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CI13
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-458-076A-6

Alignment Scores:
Pred. No.: 169 Length: 28958
Score: 92.50 Matches: 34
Percent Similarity: 41.28% Conservative: 11
Best Local Similarity: 31.19% Mismatches: 52
Query Match: 8.52% Indels: 12
DB: 1 Gaps: 4

US-09-809-545A-2 (1-203) x US-08-458-076A-6 (1-28958)
Qy 47 SerGlyProSerSerLeuValTyrrThrSerAlaMetProGlyPheProTyrrProAlaAla 66
Db 10994 GCAAGGTCTCCCTCCCACTACCCCTTCACGCGAGCGGTTCTGCTCGACGTCTCCA 10993
```


QY 1 MetThrAsnLysLeuValaValaAsnProTyrThrAsnGlyTTrpLysLeuAsnProValaVala 20
Db 535 ATGACTAATATAAAAGGCGCGTAACCCCTACACCAATGGCTGGAAATTAAATCCAGTTGTG 594

QY 21 GAlaValaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 595 GCGCGCGGTACAGCCCGCACTTCTATGACAGGACGGTGTGTGTGCCAGGCGCAACCGAG 654

QY 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 655 GAGGATCTTCCATGATACAGTGGCCCCAGTTCCTGTATATCTTCCAAATGCGCTGGC 714

QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 715 TTTCATATCCGGCGCCCACTGCTGCAGTGCATACCGAGGGGCTCACCTTGCAGGCCGT 774

QY 81 GAlaGThrValTyrAsnThrPheAlaGlyAlaAlaAlaProProProProIleProAlaTyr 100
Db 775 GGTCCGACCGGTATACAAACCTTCAAGCTGCGCGCGCCCGCCCAATCCCGGCTAT 834

QY 101 GAlaGValaValTyrGlnGluProValTyrGlyAsnLysLeuGlnGlyGlyTyrAla 120
Db 835 GCGGAGTGTGTATCAAGAGCCAGTGTATGCAATTAATGCTTACAGGGGTTCAGCT 894

QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
Db 895 GCATACCGCTACGCCCGACCCCTGCGCACTGCTGCTCCTACAGTACAGTACGGA 954

QY 141 ArgValTyrAlaAlaAspProTyrHisIleThrLeuAlaProAlaProThrTyrGlyVal 160
Db 955 CAGATTATGCTGCGCCAGCCCTTACACACACCTTGTCCACGCCCGCCCAATCCCGGCTT 1014

QY 161 GAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAla 180
Db 1015 GGTGCATATACTTTTGGCGCTTGACCGATGCCAAGATGGAGCCATGCTGATGAT 1074

QY 181 ValGlyLeuValLeuSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200
Db 1075 GTGGGCTGCTCTTCTTTCATGACAGGCTAGTATATACCAAGGGGATACACCGTTTT 1134

QY 201 AlaProTyr 203
Db 1135 GCTCCATAT 1143

RESULT 2
US-09-794-591-1
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulsec, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987) ..(1979)
US-09-794-591-1

Alignment Scores: 2,63e-85 Length: 2272
Pred. No.: 955.00 Matches: 169

Percent Similarity: 86.94% Conservative: 4
Best Local Similarity: 85.14% Mismatches: 9
Query Match: 87.94% Indels: 21
DB: 10 Gaps: 2

US-09-809-545A-2 (1-203) x US-09-794-591-1 (1-2372)

QY 1 MetThrAsnLysLeuValaValaAsnProTyrThrAsnGlyTTrpLysLeuAsnProValaVala 20
Db 1572 ATGACAAATATAAAAGGCGCGTAACCCCTTATACAAATGGCTGGAAATTGAATCCAGTTGTG 1631

QY 21 GAlaValaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 1632 GGTGACGCTTACAGTCCGCAATTTATGACAGGACGGTCTGTGTGCCAGGCGCAACGAG 1691

QY 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 1692 GAGGATCTTCCATGATACAGTGGCCCCAGTTCCTGTATATCTTCCAAATGCGCTGGC 1751

QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 1752 TTCCCGTATCCAGACGACCGCCCGCGGCGCTACCGAGGGGCGCACCTCGAGGCCGCG 1811

QY 81 GAlaGThrValTyrAsnThrPheAlaGlyAlaAlaAlaProProProProIleProAlaTyr 100
Db 1812 GGTCCGACCGGTATACAAACCTTCAAGGCGCGCGCGCCCGCCCGCAATCCCGGCTAC 1871

QY 101 GAlaGValaValTyrGlnGluProValTyrGlyAsnLysLeuGlnGlyGlyTyrAla 120
Db 1872 GCGGAGTGTGTATCAAGAGCCAGTGTATGATGCTGACAGC---ATTATGAGTGTATGC 1928

QY 120 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSer---- 138
Db 1929 TCATATACCGCTACGCCCGACGCTTACCCCTGCGCACTGCGCTCCTACAGTACAG-AAAATC 1987

QY 139 -----TyrGlyArgVa 142
Db 1988 AGTTCGCTCTGCTGACAGATGAATTTCTTGAACCTCTGCAGTTACGAGCAT 2047

QY 142 TTrpAlaAlaAspProTyrHisIleThrLeuAlaProAlaProThrTyrGlyValaVala 162
Db 2048 TTATGCTGCGGACCCCTACACACGACCTTCTCAGGCCCGCCACCTACGCGGTTGGTGC 2107

QY 162 aMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAlaVal 182
Db 2108 CATGAATGCTTTTGCACTTGTACATGATGCCAAGACTAGAGCCATGCTGATGATGCTGC 2167

QY 182 YLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPheAlaPr 202
Db 2168 TCTCGTCTTCTTTCATGACAGGCTAGTATATACCGAGGGGATACACCGTTTGTGCTCC 2227

QY 202 cTyr 203
Db 2228 ATAC 2231

RESULT 3
US-10-098-841-226
; Sequence 226, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungting
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wenman, Tom

```

; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Dimanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098, 841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pt_fl_genes Version 1.0
; SEQ ID NO 226
; LENGTH: 6930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234)..(968)
; US-10-098-841-226

Alignment Scores:
Pred. No.: 0.405 Length: 6930
Score: 108.50 Matches: 56
Percent Similarity: 35.85% Conservative: 20
Best Local Similarity: 26.42% Mismatches: 67
Query Match: 9.99% Indels: 12
DB: Gaps: 69

US-09-809-545a-2 (1-203) x US-10-098-841-226 (1-6930)
Qy 6 AlaValaAnProTyThrasnGlyTrpIysLeuAnProValaValGlyAlaValTyrSer 25
Db 336 GCCATCAATCC-----AGCTGTACCCCAATATGATCCAGTATGCT 380
Qy 26 ProAspPhe-----TyrAlaGlyThrValLeuLeuCyGlnAlaAsn 39
Db 381 CCAGAGTTTCACTTCCTGCAATTCAGCTTATGCA--ACTCTGTATGATAAGGCGCTGG 437
Qy 40 GlnGluGlySerSer-----MetTyrSerGly 48
Db 438 CCACAGAACTGCTTCTCTGTCGACGTAGGCACTTCCACCTCCAGTGCACCGG 497
Qy 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
Db 498 ACCGAGAACCAACTTACCAAGCATCTCTGCGGCTTTCAGATAT-----ACTGCG 548
Qy 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyThrValTyrAsnThrPhe 88
Db 549 GGGACACCATACAAAG-----GTCCACCGGAGCCAG 578
Qy 89 ArgAlaAlaAlaProProProProProleProAlaTyrGlyValValTyrGlnInsuPro 108
Db 579 AGTAACACTGCTCCACCCCTTACTCTCCCATCAACCCAC-----CCTATACAGCGGC 632
Qy 109 ValTyr-----GlyAsnLysLeuLeuGlnGlyGlyTyrAla 120
Db 633 ATGTATCAATCAGAAAGTCTTACCCCGACAGAAATCTGTATGCGACGAGACCTAC--- 689
Qy 121 AlaTyrArgTyrAlaGlyProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 690 -----TACACACAGCGG----- 701
Qy 141 ArgValTyrAlaAlaAspProTyr-----HisHisThrLeuAlaProAlaProThrTyr 158
Db 702 ---GCTATGCTGCCAGCTCATGTCATCCACACACACAGCGTGCACGCCAAC--- 755
Qy 159 GlyValGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAla 178

```

```

Db 756 -----ACGATCCCTCTGCTATCTATACCCAGACCTGTGCGGCCGAGAGACCAAGCT 809
Qy 179 AspAspValGlyLeuValLeuSerSerLeuGlnAla 190
Db 810 GTGGCATGGGCGATGTGTGGAGGACCAACCATGGCA 845

RESULT 4
US-09-770-445-481/c
; Sequence 481, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Jong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moesner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hubban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770, 445
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178, 472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 481
; LENGTH: 895
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-770-445-481

Alignment Scores:
Pred. No.: 0.0436 Length: 895
Score: 107.00 Matches: 52
Percent Similarity: 36.60% Conservative: 19
Best Local Similarity: 26.80% Mismatches: 63
Query Match: 9.85% Indels: 60
DB: Gaps: 11

US-09-809-545a-2 (1-203) x US-09-770-445-481 (1-895)
Qy 42 GlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGlyPhe 61
Db 869 GGTATGTCCTTACCAAGTCCA-----ACACACAGCAGCAACCGTCATAT 825
Qy 62 ---ProTyrProAlaAla-----ThrAlaAlaAlaTyrArgGlyAla 75
Db 824 GATTCAACACAGAGCAGCATGATGTTCAACTCCAGCAGACCATCTGATGTGTCA 765
Qy 76 HisLeuArgGlyArg-----GlyArgThrValTyr 85
Db 764 AACATGGCTCAACACAGCAATATGTTATCATCACTAGTGGCCCTACTCAGAACTTAT 705
Qy 86 AsnThrPheArgAlaAlaAlaPro-----ProProProle 97
Db 704 CCTTCATATAGCTGTGCGACGACCATCTGATGTTATATGTGACACACACGACGACTT 645
Qy 98 ---ProAlaTyr-----GlyGlyValValTyrGlnInsuPro 108

```

```

Db 644 GCCCAGCTTATGAGCAGCAGCTGCTCAGCCAGCTTCTGCTGTCGACAACTTCAAGT 585
Qy 109 VALTYGTYAAsnlyeuleuenginglyTYrAlaAlaTYrArgTYrAlaInProth 128
Db 564 GGGTACGGGCAAGTACCTCCACCGGTGGCTATAGTTCGATCCCTCCACAGCGCGGT 525
Qy 129 ProAlaThraAlaAlaTYrSerAep---SerTYrGlyArgValTYrAlaAlaAepPro 147
Db 524 TATGTAATACCCCGCTCAAGCAATGAACTATGGA-----TACATTGGCTCTCAG 471
Qy 148 TYrHisHleThLeuAlaProAlaProThrTYrGlyValGlyAlaMetAAsnAlaPheAla 167
Db 470 TAT-----CCTAGCTATGAGGTGGAACCATCAGCATATGCT 432
Qy 168 ProLeuThraPheAlaTYrThraArgSerHleAlaAAspValGlyLeuValLeuSerSer 187
Db 431 GACCTACCTGGCCAAACCGCTTATTC-----405
Qy 188 LeuGlnAlaSerIleTYrGlnGlyTYrAAsnArgPheAla 201
Db 404 ---CAGACTGCACCTCTCAGCGCGGTATGACAAATCAGCA 366

```

RESULT 5

```

US-09-996-634-62
; Sequence 62, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 821
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(821)
; OTHER INFORMATION: n is a, c, g, or t/u.
US-09-996-634-62

```

Alignment Scores:

```

Pred. No.: 0.27 Length: 821
Score: 96.50 Matches: 59
Percent Similarity: 28.46% Conservative: 15
Best Local Similarity: 22.69% Mismatches: 95
Query Match: 9.07% Indels: 91
DB: Gaps: 9

```

US-09-809-545A-2 (1-203) x US-09-996-634-62 (1-821)

```

Qy 1 MetThraAsnlyeAlaValaAsnProTYrThraAsnGlyTYrlyeLeuAAsnProVal 20
Db 70 ATGGCGGAGCAGCTGCTGGCCGAGCCGAGCCCAATGCCGCAATGCTGCACACCGATTCCG 129
Qy 21 GtAlaAlaValTYrSerProAAspPheTYrAlaGlyThraValLeuLeuGlyGlnAlaAAsn 40
Db 130 GCGCAGCGCTTCGACCGACGACCGCTGCGGTATCACTCCGTCGCTTCAAAACCG 189
Qy 41 GluGlySerMetTYrSerGlyProSer-----SerLeuValTYr 54

```

```

Db 190 AGGCGCTGGCGCAGAGACTCAAGTCCGACCCGGTGCTCCAAACCCGGGCTGGTCAAC 249
Qy 55 ThrSerAlaMetProGlyPheProTYrProAlaAlaThraAlaAlaTYrArgGly 74
Db 250 TCCGATGGCTGCGCCCAACAAACCAACCGCGCATCAGCATCCGCGGGGACACCGCGGA 309
Qy 74 -----74
Db 310 GGGAGGGCCCGGANTCGGATNCAAGGGTTCGACCGCGCTGCGCTTCNGATTGGA 369
Qy 75 -----AlaHisLeuArgGlyArg-----GlyArgThraValTYr 85
Db 370 CCCGACAGCTACCCCGGTATGGCAGCTACGGGAGAAACAACCTGGCCGACCGCCAC 429
Qy 86 AsnThraPheArgAlaAlaAlaProProProIleProAlaTYrGlyValValTYr 105
Db 430 CTCGGCTGTGACAGTTATCCGCCCGCAGCGCGGACCGGACCGCGCTGTGTGT 480
Qy 106 GlnGluProValTYrGlyAsnlys-----LeuLeuGlnGlyGly-----118
Db 481 -----GGTTTCGGCGCGCGCGCATCTGTCTTCAAGAGAGACGGCATTTGATCTTA 534
Qy 119 -----TyrAlaAlaTYrArgTYrAlaGlnProThraProAlaThr 131
Db 535 CGGCCAGTCTCCGAAACTGCAAGTGGGGGTCAACCGCGGACGGCCGATCCAGCACT 594
Qy 132 AlaAlaAlaTYrSerAAspSerTYr-----139
Db 595 GGGGAGGATTTCCGATCGATCGACCGACCCGCGTGGCGCAATCTCGGTTCC 654
Qy 140 -----GlyArgValTYrAlaAlaAAspProThraHis-----150
Db 655 GCTGGCTGGGCGCGCGGAGCGGACGCTGGCGG-----CATGTGCG 699
Qy 151 -----ThraAlaProAlaProThraTYrGly 159
Db 700 CATATACCCGAACCTGAGCGCTGAGCAATGTTCCCTTACCCCGCGGTTCCGCT 759
Qy 160 ValGlyAlaMetAsnAlaPheAlaProLeuThraPheAlaTYrThraArgSerHisAlaAAsp 179
Db 760 GCTGGAATCTGCAAGCGGTGATCGGTCAGCAGCACCGGTTGATGATGATCGCGAC 819

```

RESULT 6

```

US-09-815-242-4035/c
; Sequence 4035, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308

```

PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4035
LENGTH: 2658
TYPE: DNA
ORGANISM: *Pseudomonas aeruginosa*
US-09-815-242-4035

Alignment Scores:
Pred. No.: 148 Length: 2658
Score: 97.50 Matches: 60
Percent Similarity: 36.19% Conservative: 16
Best Local Similarity: 28.57% Mismatches: 74
Query Match: 8.98% Indels: 61
Gaps: 10

US-09-809-545a-2 (1-203) x US-09-815-242-4035 (1-2658)

Qy 14 TrpLysLeuAnProVala1G1yAlaValTyrSerProAspPheTyrAlaG1yThrVal 33
Db 2263 TGGCAGCATGCGCTGCGACCGGGTACGGCGCAGCACC-----CGGCGACGGCTTCA 2210

Qy 34 LeuLeuCysGln-----Ala 38
Db 2209 GCGGCTGCGCAACGATGTCGCCCGCGCACCCAGTCCGCGCCAGCTTCAAGCCACCGT 2150

Qy 39 AsnGlnGluGlySer-----SerMetTyrSerGlyProSerSerLeuValTyr 54
Db 2149 GCGCGAGCGGGGTGCTGTGACAGCAGTCTTGATATACGGGCTCGCTCGCGCTCGT 2090

Qy 55 ThrSerAlaMetPro-----GlyPheProTyrProAla--- 65
Db 2089 GCGGCGCTCTTCCAGCAGTTCGCGACGATCCGCGGGGATGCGCTCCACCCAGCGCA 2030

Qy 66 -----AlaThrAlaAlaAlaTyrArgGlyAlaHis 76
Db 2029 GCAGGCTGCGATGAGAGCGCGCATGCGCGGTGCGCGGTGCGCGAGTGTGGZACACCG 1970

Qy 77 LeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProPro--- 95
Db 1969 AGCGGAGCAAGCGCTGCGCGAGTTCCTCGCTCGCGCGGCGCGCGCTCCAGGT 1910

Qy 96 -----ProIleProAlaTyrGlyValValTyrGlnProValTyr 110
Db 1909 CCTGCGCGCAGCGAGCG 1862

Qy 111 GlyAsnLysLeuLeuGlnGly-----GlyTyrAlaAlaTyrArgTyrAlaGlnProThr 128
Db 1861 GGGCGCGAGAGCGCGCGCGCGCTCGGGGCTCAGCGCGCTGCGCGCGCGCGCGCGCGCG 1803

Qy 129 ProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspProTyr 148
Db 1802 CCGAGCGCGCGAGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1764

Qy 149 HisHisThrLeuAlaProAlaProAlaProThrTyrGlyVal-----GlyAlaMetAsnAlaPhe 166
Db 1763 CACCAACCAACCGCTCTCCGGGCGAGCGTGCACGACGAGCGCGCGCGCGCGCGCGCG 1704

Qy 167 AlaProLeuThrAspAlaLysThrArgSer 176
Db 1703 GCCCAGCGCGCGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1674

RESULT 7
US-09-712-363-143
Sequence 143, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS

FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 143
LENGTH: 3285
TYPE: DNA
ORGANISM: *Mycobacterium tuberculosis*
US-09-712-363-143

Alignment Scores:
Pred. No.: 1.93 Length: 3285
Score: 97.50 Matches: 66
Percent Similarity: 30.91% Conservative: 19
Best Local Similarity: 24.00% Mismatches: 95
Query Match: 8.98% Indels: 96
Gaps: 12

US-09-809-545a-2 (1-203) x US-09-712-363-143 (1-3285)

Qy 1 MethrAsnLysLysAlaValAsnProTyrThrAsnGlyTyrLysLeuAnProVala1 20
Db 2203 ATGGCGAGCGACGTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2262

Qy 21 GlyAlaValTyrSerProAspPheTyrAlaG1yThrValLeuLeuCySGlnAlaAsnGln 40
Db 2263 GCGCAGCGCTTCCAGCAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2322

Qy 41 Glu-----GlySerSerMetTyrSerGlyPro-----SerSerLeuValTyr 54
Db 2323 GAGGCGGTGCGGCGAGCACTCAAGTCCGACCGCGGTGCTCCAAACCGCGCGGTGTCMAC 2382

Qy 55 ThrSerAlaMetProGlyPheProTyrProAlaThrAlaAlaAlaAlaTyrArgGly 74
Db 2383 TCCATGCGTGCAGCAACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2442

Qy 75 AlaHisLeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaPro--- 93
Db 2443 -----GGAGAGCGCGCGGTGCGGATCAAGCGGTGCGAGCGCGCGCGCGCGCGCG 2490

Qy 94 -----ProProIleProAlaTyrGly----- 101
Db 2491 GAATTGACCGCGACGTAACCGGTGATGGCAGCTACGGGAGAACCAACCTGCGCGCC 2550

Qy 102 -----GlyValValTyrGlnGluPro----- 108
Db 2551 ACGCCACCTCGCGCTGTGATCAAGTACCGCCCGCGACCGCGACCGCGCGCTGTGTG 2610

Qy 109 -----ValTyrGlyAsn 112
Db 2611 GTTTCGCGCGCGCGCGCATCTGATCTTACAGAGAGAGCGCGCATTTTATCTAGGCGAG 2670

Qy 113 LysLeu-----LeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
Db 113 LysLeu-----LeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131

Db	2671	TCCTGAAACTGCAGTG-666-----CGTACCGGCGCGAGCGCGATCAGCACT	2723
Qy	132	AlAlAlAlaIYrSsrrAaPseTyr-----	139
Db	2724	GGGGCGAGGTATTTCGATCGACATCGACCGCAACCGCGTGGCGAGACTCGCGTTCC	2783
Qy	140	-----GIYrGValTYrAlAlaAaPProTYrHisA-----	150
Db	2784	GCTGGCTGCGCGCGCGCGAGCGGACGTGGCGC-----CATGTGCG	2828
Qy	151	-----ThIleuAlAProhlaAProhThyGly	159
Db	2829	CTATGACCCCACTGAGCCCTGAGCATGTCGCTTCACCCCGCGCGGCGTTCCGT	2888
Qy	160	ValGIYAlaMeAaAlaAlaPheAlaProIeuThrAspAlaIYrThrArgSerHisAlaAaP	179
Db	2889	GCTGGAAATCTGTGACGGGTGATCGGTCAGACGACACCGGTGTGATGAGCATGCGAC	2948
Qy	180	AaPValGIYIleuValIeuSerSerIeuAlaIYrSerIYrGln	194
Db	2949	CGCAGCCAACTT-----CCCTTCACAGCAGACCGTTTCCGA	2994

```

RESULT 8
US-08-754-311B-1
; Sequence 1, Application US/08754311B
; Patent No. US20020004221A1
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzert, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONISTS
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbe
; STREET: 4 Embarradero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/754,311B
; FILING DATE: 21-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/195,152
; FILING DATE: 14-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccarlin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3231 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
;
US-08-754-311B-1

Alignment Scores:
Pred. No.: 2.66 Length: 3261
Score: 96.00 Matches: 45
Percent Similarity: 37.57% Conservative: 23
Best Local Similarity: 24.86 Mismatches: 63

```

Query Match:	8.84%	Indels:	51
DB:	8	Gaps:	6

US-09-809-545A-2 (1-203) X US-08-754-311B-1 (1-3231)

Qy 42 GlySerMetTyr-----SerGlyProSerSerLeu 52

Db 1062 GGATCCAATTGTACGGCTGCAGCTCGGCCAGCAATCCGCTGGACGGAGGACAGTGGCG 1121

53 Val Tyr His Ser Ala Met Pro Gly Phe Pro Tyr Pro Ala Ala Ile Ala Ala Ala Ile /2

[illegible]

α_1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

113 IValLeuIleuG]nG]vG]vTvrA]aA]atVrArGtVrA]aG]nProThrProA]aThrAla 132

Db 1272 GCCAAGATGGAACCCGGAGCGGCGAGCTGCGGCGGCTGCCTACTTGACGCCCACTATGCC 1331

QY	133 A1aA1a-----	134
----	-----------------	-----

Db 1332 GCCAGCGGCAACAACACTCGCAGCTGTACAGCAGTCCGTACGCCGGCTACACAACCTTC 1391

QY 135 -----TyrSerAspSerTyrGlyArgValTyrAlaAlaAsp 146

Db 1392 GGGCAGCAGGACTACGGCGGCTACTACAACGAGCAGTACGGCACTATTACAGTCCGCC 1451

14 / PROLYTHISINLEUADPLOWADPO-----IMLYGVVADGLY--- 161

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

$\frac{1}{2}$
 $\frac{1}{3}$
 $\frac{1}{4}$
 $\frac{1}{5}$
 $\frac{1}{6}$
 $\frac{1}{7}$
 $\frac{1}{8}$
 $\frac{1}{9}$
 $\frac{1}{10}$
 $\frac{1}{11}$
 $\frac{1}{12}$
 $\frac{1}{13}$
 $\frac{1}{14}$
 $\frac{1}{15}$
 $\frac{1}{16}$
 $\frac{1}{17}$
 $\frac{1}{18}$
 $\frac{1}{19}$
 $\frac{1}{20}$
 $\frac{1}{21}$
 $\frac{1}{22}$
 $\frac{1}{23}$
 $\frac{1}{24}$
 $\frac{1}{25}$
 $\frac{1}{26}$
 $\frac{1}{27}$
 $\frac{1}{28}$
 $\frac{1}{29}$
 $\frac{1}{30}$
 $\frac{1}{31}$
 $\frac{1}{32}$
 $\frac{1}{33}$
 $\frac{1}{34}$
 $\frac{1}{35}$
 $\frac{1}{36}$
 $\frac{1}{37}$
 $\frac{1}{38}$
 $\frac{1}{39}$
 $\frac{1}{40}$
 $\frac{1}{41}$
 $\frac{1}{42}$
 $\frac{1}{43}$
 $\frac{1}{44}$
 $\frac{1}{45}$
 $\frac{1}{46}$
 $\frac{1}{47}$
 $\frac{1}{48}$
 $\frac{1}{49}$
 $\frac{1}{50}$
 $\frac{1}{51}$
 $\frac{1}{52}$
 $\frac{1}{53}$
 $\frac{1}{54}$
 $\frac{1}{55}$
 $\frac{1}{56}$
 $\frac{1}{57}$
 $\frac{1}{58}$
 $\frac{1}{59}$
 $\frac{1}{60}$
 $\frac{1}{61}$
 $\frac{1}{62}$
 $\frac{1}{63}$
 $\frac{1}{64}$
 $\frac{1}{65}$
 $\frac{1}{66}$
 $\frac{1}{67}$
 $\frac{1}{68}$
 $\frac{1}{69}$
 $\frac{1}{70}$
 $\frac{1}{71}$
 $\frac{1}{72}$
 $\frac{1}{73}$
 $\frac{1}{74}$
 $\frac{1}{75}$
 $\frac{1}{76}$
 $\frac{1}{77}$
 $\frac{1}{78}$
 $\frac{1}{79}$
 $\frac{1}{80}$
 $\frac{1}{81}$
 $\frac{1}{82}$
 $\frac{1}{83}$
 $\frac{1}{84}$
 $\frac{1}{85}$
 $\frac{1}{86}$
 $\frac{1}{87}$
 $\frac{1}{88}$
 $\frac{1}{89}$
 $\frac{1}{90}$
 $\frac{1}{91}$
 $\frac{1}{92}$
 $\frac{1}{93}$
 $\frac{1}{94}$
 $\frac{1}{95}$
 $\frac{1}{96}$
 $\frac{1}{97}$
 $\frac{1}{98}$
 $\frac{1}{99}$
 $\frac{1}{100}$

RESULT 9

US-09-9/4-300-1655
: Sequence 1655, Application US/09974300

; GENERAL INFORMATION:

APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION:	Expression
TTT TTTTTTTT 1000F EOO ITT	

CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05

PRIOR FILING DATE: 2000-10-06

; PRIOR AFFIDAVIT NUMBER: 00/277,220
 ; PRIOR FILING DATE: 2001-03-27
 ;

```

; SOFTWARE: FastSeq for Windows Version 4.0

```

LENGTH: 657

ORGANISM: *Bacillus licheniformis*

```

; NAME/KEY: misc_feature
LOCATION: (1)_(657)

```

OTHER INFORMATION: II = A, I, C OL G
US-09-974-300-1655

Alignment Scores:

Pred. No.: 0.569 Length: 657
 Score: 94.00 Matches: 38
 Percent Similarity: 40.15% Conservative: 17
 Best Local Similarity: 27.74% Mismatches: 58
 Query Match: 8.65% Indels: 24
 DB: 10 Gaps: 8

US-09-809-545a-2 (1-203) x US-09-974-100-1655 (1-657)

Qy 2 ThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTPbLysLeuAsnProValValGly 21
 Db 100 ACAAAACCAAAACCAATTAACCA--ACAGGGGCTTAAACGAACGAGTAACTCC 156
 Qy 22 -----AlaValTyrSerProAspPhe--TyrAlaGlyThrValLeuLeu 35
 Db 157 TTAACGGCCCCCGGGCTTTTCCAAGATTTCTTATACAAAACCTGGCTCTCTT 216
 Qy 36 CysGlnAlaAsnGlnGlySerSerMetTyrSerGlyProSerSerLeuValTyrThr 55
 Db 217 ACCACCAAAACCAAGGTCGGGTTTC-----CCCTTAAGGAAATTTAAC 267
 Qy 56 SerAlaMetProGlyPheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAla 75
 Db 268 CCCTTTAAACCAATTTCACTTTTAACTTAAACGCAACCCCAATCCCAACCCCTC 327
 Qy 76 HisLeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProPro 95
 Db 328 CATATTCATCCTTGTTGTT-----TTCAAAATTTTTCATTCATTAACCCCTCG 378
 Qy 96 ProIleProAla-----TyrGlyGlyValValTyrGlnGly 107
 Db 379 AAAAACCCCTCCCGGTACAACTATTTTCAATACAAACGGGGCTTGCGCTCGGAA 438
 Qy 108 ProValTyrGlyAsnLysLeuLeuGlnGly-----TyrAlaAlaTyr 122
 Db 439 CCAATTAACCCCAAC--CTGGCCAAAGCGCGCGGTTTACTCCCTTAC 486

RESULT 10

US-09-822-830A-410/C
 ; Sequence 410, Application US/09822830A
 ; Patent No. US20020142952A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genetics Institute, Inc.
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fechtel, Kim
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Howes, Steven H.
 ; APPLICANT: Resnick, Richard J.
 ; APPLICANT: Gulukota, Kamalakkar
 ; APPLICANT: Graham, James R.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
 ; FILE REFERENCE: GIN 6402
 ; CURRENT APPLICATION NUMBER: US/09/822, 830A
 ; CURRENT FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: 60/195,604
 ; PRIOR FILING DATE: 2000-04-06
 ; NUMBER OF SEQ ID NOS: 631
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 410
 ; LENGTH: 2273
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-822-830A-410

Alignment Scores:

Pred. No.: 3.8 Length: 2273
 Score: 92.50 Matches: 48
 Percent Similarity: 33.72% Conservative: 10
 Best Local Similarity: 27.91% Mismatches: 73
 Query Match: 8.52% Indels: 41
 DB: 10 Gaps: 9

US-09-809-545a-2 (1-203) x US-09-822-830A-410 (1-2273)

Qy 29 TyrAlaGlyThrValLeuLeuLeuGlnAla---AsnGlnGlySerSerMetTyrSer 47
 Db 2182 TATGCACAGACACC-----CAGGCATATGGCAACCAAGCTATGAACTATGCA 2132
 Qy 48 GlyProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThr 67
 Db 2131 CAGCCCACTGATGTCAGCTATACCCAGGCTCAGACCACTGCAACCTATGGCAGACGCC 2072
 Qy 68 AlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyAlaGlnThrValTyrAsnThr 87
 Db 2071 TATCAACTTCTTAT-----GGACAGCTCCCACTGCTTAATACTACT 2030
 Qy 88 PheArgAlaAlaAlaProProProIleProAlaTyrGlyGlyValValValTyrGlnGly 107
 Db 2029 -----CNACTGCCCCCCAG-----GCATACAGCCAG 2003
 Qy 108 ProValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnPro 127
 Db 2002 CCTGTCAGGGGTAT-----GGCACTGGTGCTTATGATACCAACCACTGCT 1958
 Qy 128 ThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro 147
 Db 1957 ACAGTCAACCAACCCAGGCC-----TCTATGCACTCAGTCT 1919
 Qy 148 TyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAla 167
 Db 1918 GCATATGGCACTCAGCTCTTATTCAGCTATAGGAG-----CAG 1877
 Qy 168 ProLeuThrAspAlaLysThrArgSerHisAlaAspValGlyLeuValLeuSerSer 187
 Db 1876 CCAGAGCACTGACACCTACAGACGAGATGAAACCAAGCCACATGACACTAGTCAA 1817
 Qy 188 LeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArg 199
 Db 1816 CCTCAATCTAC-----ACAGGGGGGTACCAACAG 1787

RESULT 11

US-09-880-107-3769
 ; Sequence 3769, Application US/09880107
 ; Patent No. US20020142981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horne, Darci T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scherf, Uwe
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-NO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3769
 ; LENGTH: 2390
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X66899
 ; US-09-880-107-3769

Alignment Scores:

Pred. No.: 4.05 Length: 2390
 Score: 92.50 Matches: 48
 Percent Similarity: 33.72% Conservative: 10
 Best Local Similarity: 27.91% Mismatches: 73
 Query Match: 8.52% Indels: 41
 DB: 10 Gaps: 9

US-09-809-545A-2 (1-203) x US-09-880-107-3769 (1-2390)

Qy 29 Trralaglythrvalleuencysginala---Asnglglucllysermettyrser 47
Db 128 TATGCACACACCC-----CAGGCTATGGCCACMAAGATATGAGACTTGA 178	
Qy 48 GlyProSerSerleuValtyrThrSerAlaMetProGlyPheProtyrProAlaIatThr 67	
Db 179 CAGCCCATGATGTAGCTATACCGAGCTCAGACCATGCAACCTATGGCAGACCGCC 238	
Qy 68 AlaAlaAlaAlaAlaTyArgGlyAlaHisLeuArgGlyAlaArgGlyAlaArgGlyAlaArgThr 87	
Db 239 TATGCACCTCTTAT-----GGACAGCTCCCTCAGTGTATATCTACT 280	
Qy 88 PheArgAlaAlaAlaAlaProProProIleProAlaTyArgGlyAlaValAlaValtyrGlnGlu 107	
Db 281 -----CCTACTGCCCTCCAG-----GCATACAGCCAG 307	
Qy 108 ProValtyrGlyAsnlyLeuLeuGlnGlytyrAlaAlaTyArgGlyAlaGlnPro 127	
Db 308 CCTGTCCAGGGGAT-----GGCAGTGTGCTTATGATACACCATGTCT 352	
Qy 128 ThrProAlaThrAlaAlaAlaTySerAspSerTyArgGlyAlaValtyrAlaAlaAspPro 147	
Db 353 ACAGTCACCCACCCAGCGC-----TCCATATGAGCTCAGTCT 391	
Qy 148 TyrHisHisThrLeuAlaProAlaProThrTyArgGlyValGlyAlaMetAsnAlaPheAla 167	
Db 392 GCATATGGCATCAGCTGCTTATCCAGCTATGGCAG-----CAG 433	
Qy 168 ProLeuThrAspAlaAlaTyThrArgSerHisAlaAspAspValGlyLeuValLeuSerSer 187	
Db 434 CCAGCAGCCACCTGACCTTACAGACCGCAGATGAAACAGCCCATGAGTACTCAA 493	
Qy 188 LeuGlnAlaSerIleTyArgGlnGlytyrAsnArg 199	
Db 494 CCTCATCTAGC-----ACAGGGGGTTCACACAG 523	

RESULT 12

US-09-904-968A-1

/ Sequence 1, Application US/09904968A

/ Publication No. US20030008288A1

/ GENERAL INFORMATION:

/ APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE

/ APPLICANT: GERMINO, Gregory

/ APPLICANT: MATNICK, Terry

/ APPLICANT: PHADEEKITCHARON, Bunyong

/ TITLE OF INVENTION: DETECTION AND TREATMENT OF POLYCYSTIC KIDNEY DISEASE

/ FILE REFERENCE: JHU1680-2

/ CURRENT APPLICATION NUMBER: US/09/904,968A

/ CURRENT FILING DATE: 2001-07-13

/ PRIOR APPLICATION NUMBER: US 60/283,691

/ PRIOR FILING DATE: 2001-07-13

/ PRIOR APPLICATION NUMBER: US 60/218,261

/ NUMBER OF SEQ ID NOS: 113

/ SOFTWARE: PatentIn version 3.0

/ SEQ ID NO: 1

/ LENGTH: 53522

/ TYPE: DNA

/ ORGANISM: Homo sapiens

US-09-904-968A-1

Alignment Scores:

Pred. No.: 224 Length: 53522

Score: 92.00 Matches: 58

Percent Similarity: 34.60% Conservative: 24

Best Local Similarity: 24.47% Mismatches: 83

Query Match: 8.47% Indels: 72

DB: 9 Gaps: 11

US-09-809-545A-2 (1-203) x US-09-904-968A-1 (1-53522)

Qy 18 ProValGlyAlaValAlaTy-----SerProAspPheTyArgGlyThrValLeu 34
Db 20809 CCTCTTACGTGAGAGATATCCCTGCTCCTGACACAGCTCAGCCCTG--- 20865	
Qy 35 LeuCyGlnAlaAsnGlnGluIlySerSerMetTySerGlyProSerSerleuValTy 54	
Db 20866 -----GCACAGTGTCTTT 20880	
Qy 55 ThrSerAlaMetProGlyPheProtyrProAlaAla----- 66	
Db 20881 TCAGCTGCCACGAGAGGCTGCTTACAGCCAGAGGCTGAGGCTTCTTCACCC 20940	
Qy 67 -----ThrAlaAlaAlaAlaTyArgGlyAlaHisLeuArgGlyAlaArgGlyAlaArgThr 83	
Db 20941 GCCACGAGGCTCCGAGCCCTTCGAGAGAGGCTGTGCTGTGGGGCGGCCAGCCC 21000	
Qy 84 ValTyAsnThrPheArg-----AlaAlaAlaProProProIlePro 98	
Db 21001 TCCAGTCCCTCTTGTCCCTGCTGCTCCTGCTCCGCCCCCGCACCTCTGCCCC 21060	
Qy 99 AlaTyArgGlyValAlaValtyrGlnGluProValTy----- 110	
Db 21061 ACCTGTAGGGGCCCAACCTCCTCCAGACAGCTTCCCTGCCCAAGGGCCACCTG 21120	
Qy 111 -----GlyAsnlyLeuLeuGlnGlytyrAlaAlaTyArgGlyAlaGlnPro 127	
Db 21121 GTGGGGCCCCACGACCTGTGCTGTGCGCAGTACAGCCTTCACATCCTGCCCG 21180	
Qy 128 ThrProAlaThrAlaAlaAla-----TySerAspSerTyArgGlyAlaTyArgAlaAla 145	
Db 21181 CTCCTGTACCTGCCACACAGCTGGACCTTGGAGACGGCTCCCGAGGTGATGCCGT 21240	
Qy 146 AspPro-----TyHisHisThrLeuAlaPro 154	
Db 21241 GGGCCGGCTCCTCGCATGTGCTGCTGGGGCGCTATCAGTACGGCCGCTG 21300	
Qy 155 AlaProThrTyArgGlyAlaValAlaMetAsnAlaPheAlaProLeuThrAspAlaTyThr 174	
Db 21301 GCC-----CTGGGGCGCGCTCAGCCCTGCTGGG---ACAGACGTGACAGTG 21345	
Qy 175 ArgSerHisAlaAspAspValGlyLeuValLeu-----SerSer 187	
Db 21346 GAGCGGCACCTGCCCGCTGAGCTCGTGTCCCTGCTCGGTGACAGTACAGAGGC 21405	
Qy 188 LeuGlnAlaSerIleTy-----GlnGlyGlytyrAsnArgPheAlaProTy 203	
Db 21406 CTCGACCTCAGCATCCAGAACCGGCTGTTCAGGCTGTGAGGCGGCTTAC 21456	

RESULT 13

US-10-125-635A-452

/ Sequence 452, Application US/10125635A

/ Publication No. US20030039635A1

/ GENERAL INFORMATION:

/ APPLICANT: Galiger, Alexander

/ APPLICANT: Smithgall, Molly D.

/ APPLICANT: Carter, Darlick

/ APPLICANT: Cheever, Martin A.

/ APPLICANT: McNeill, Patricia D.

/ APPLICANT: Sutherland, R. Alec

/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR WTI

/ FILE REFERENCE: 210121.465C7

/ CURRENT APPLICATION NUMBER: US/10/125,635A

/ CURRENT FILING DATE: 2002-07-19

/ NUMBER OF SEQ ID NOS: 461

/ SOFTWARE: FastSeq for Windows Version 3.0

/ SEQ ID NO: 452

/ LENGTH: 969

/ TYPE: DNA

/ ORGANISM: Homo sapiens

US-10-125-635A-452

Alignment Scores:	
Pred. No.:	1.64
Score:	91.50
Percent Similarity:	37.68%
Best Local Similarity:	25.12%
Query Match:	8.43%
DB:	9
length:	965
Matches:	52
Conservative:	26
Mismatches:	77
Indels:	32
Gaps:	8

US-09-809-545A-2 (1-203) X US-10-125-635A-452 (1-969)

Oy ProTyrThrAsnGlyTyrTrpLysLeuAsnProValGlyAlaValTyrSerProAspPhe 28
 Db 42 CTTGGCTACAGCTGGGTGGCTT-----GACCGTTGGCTGG 74
 Oy TyrAlaGlyTyrThrValLeuLeuCysGlnAlaAsnGlnGlySerSerMetTyr----- 46
 Db 75 TATGCTGGGTCCGTC-----TCTGGTACCCCGGCTGGTGTACCGGTGC 119
 Oy 47 -----SerGlyProSerSerLeuValTyrThrSerAlaMetProGlyPhe 61
 Db 120 TCACGGCTCCGACGTTCCGGGACCTTGAAACGACACTGTGCGGACAGTCCGCTCGGGTGG 179
 Oy ProTyrProAlaAlaThrAla-----AlaAlaAlaTyrArgGlyAlaAlaLeu 77
 Db 180 TGGTGTGGGTCCGCACTGGCCGGTTACGGGTGCACACAGTGGGCTCCGGTTCTGACTT 239
 Oy 78 ArgGlyArg-GlyArgThrValTyrAsnThrPheArgAla-----AlaAla 92
 Db 240 CGCACCGCGGGGTGATCGGATACGATTACGGTTCCGTGGGTGGTCCGGACCGCGCGGACCC 299
 Oy 92 AProProProProlIleProAlaTyrGlyGlyValTyrGlnGlnProValTyrGly-- 111
 Db 300 GCGCGCGCGCGCGCGCGCGCGCGCACTCTTCATCAAAACAGAACCGACTGGGGTGG 358
 Oy 112 -----AsnLysLeuLe 115
 Db 360 TGCAGAACCGACGGAAGAACACTGCTTACGGCATTCACCGTTCACTTCTCGGCGCACTT 419
 Oy 115 GlnGlnGlyTyrTyrAlaAlaTyrArgTyrGlyAlaGlnProThrProAlaThrAlaAlaTyr 135
 Db 420 CACTGGCACAGCGGAGGCTGTGCTTACGGGCGCTTGGGTCTCTCTCGCGCCACCGAGGC 479
 Oy 135 rSerAspSerTyrGlyArgValTyrAlaAlaAspProTyr---HisThrLeu-Alap 155
 Db 480 GTCAATCCGCGCAGGCGACAGATGTTCTTAAGCGCGCTTACCTGGCCACGTCCTTCGAGAG 539
 Oy 154 roAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAlaProLeuThr-AspAlaLys 173
 Db 540 CCAGCCCGCTATTCGCAATC--AGGTTACAGACGGGTCACTTCGACGGGAGACCGCAG 596
 Oy 174 ThrArgSerHisAla 178
 Db 597 CTACGGTCAACAGCC 611
 RESULT 14
 US-09-887-576-788
 ; Sequence 788, Application US/09887576
 ; Patent No. US20020144047A1
 GENERAL INFORMATION:
 ; APPLICANT: Budworth, P.
 ; APPLICANT: Brown, D.
 ; APPLICANT: Chang, H.
 ; APPLICANT: Zhu, T.
 ; APPLICANT: Han, B.
 ; APPLICANT: Wang, X.
 ; APPLICANT: Cooper, Bret
 TITLE OF INVENTION: Promoters for regulation of plant expression
 FILE REFERENCE: 1360.001US1
 CURRENT APPLICATION NUMBER: US/09/887,576
 CURRENT FILING DATE: 2001-06-25
 PRIOR APPLICATION NUMBER: US 60/213,848
 PRIOR FILING DATE: 2000-06-23
 PRIOR APPLICATION NUMBER: US 60/214,087

```

: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 60/258,692
: PRIOR FILING DATE: 2000-12-29
: NUMBER OF SEQ ID NOS: 875
: SOFTWARE: FastSeq for Windows Version 4.0.
: SEQ ID NO 788
: LENGTH: 1743
: TYPE: DNA
: ORGANISM: Oryza sativa
:
: US-03-887-576-788

```

Pairwise Similarity:	0.95	1.743
Pred. No.:	4.29	Length:
Score:	90.50	Matches:
Percent Similarity:	31.46%	Conservative:
Best Local Similarity:	23.66%	Mismatches:
Query Match:	8.33%	Indels:
DB:	10	Gaps:
		7

US-09-809-545A-2 (1-203) X US-09-887-576-788 (1-1743)

Oy 37 GlnAlaIngIngluInglySerSerMetTySerSerGlyProSerSerLeuValTyrThySer 56
 ||| ::: ::: ::: ||||| |||||
 Db 1189 CAAACCTTCACACTCCTGCCTGCTCGAGGCCAACCCCTGGCTATTTCACACCACCACTTAAGC 1248
 Oy 57 AlaMetProGlyPheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHis 76
 ||| ||| ::: ||| ::: |||
 Db 1249 GCTGCTTCACACTACCCGGGTGCAGAGGGTCTGCTCATGCTGTATAT----- 1296
 Oy 77 LeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProPro-- 95
 ||| ||| ||| ||| ||| |||
 Db 1297 -----GGGGCGGCCACCAACACAG 1314
 Oy 96 -----ProIleProAlaTyrGlyGlyValValTyrGlu 107
 ||| ||| |||
 Db 1315 ACTGSGTAGTGATCCAGCCCCAACAAGGAGGATATATGTCAGGGCACCTATGGCGCA 1374
 Oy 108 ProValTyrGlyAnLysLeu----- 114
 ||| ||| ||| |||
 Db 1375 CCTCGCAGGAGCAAGAAAGGCTCCTCTACACTTCTCTTACGGACAGCGCGCCTCCT 1444
 Oy 115 -----LeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGln----- 126
 ||| ||| ||| ||| |||
 Db 1435 GGATCTGCTCCTGGTGGGTATGGCCAGATATGTTACTCTAGAACAAGAGCTATGGC 1494
 Oy 127 -----ProThrProAlaThrAlaAlaAlaTyrSerSerSerTyrGlyArgValTyr 143
 ||| ||| ||| ||| ||| |||
 Db 1495 GCACCTCCACCTTACCTGTGGTGCACCTGTGCAGGCCACAGGCTATGTATGCACAG 1554
 Oy 144 Ala---AlaAspProTyrTrpHisThrLeuAlaProAlaProThrTyrGlyValGlyAla 162
 ||| ||| ||| ||| ||| ||| |||
 Db 1555 TCCATATGCTATCCTTAT-----GGCAGTGAACCTACGGGACCTGGCC 1599
 Oy 163 MetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspASP 180
 ||| ||| ||| ||| ||| |||
 Db 1600 -----GCATATTCTACTGAAGCTACAACTGCTGCTGCCACAGAC 1641

RESULT 15 US-09-938-842A-1143
 Sequence 1143, Application US/09938842A
 Patent No. US20020160378A1
 GENERAL INFORMATION:
 APPLICANT: Harper, Jeff
 APPLICANT: Kreps, Joel
 APPLICANT: Wang, Xun
 APPLICANT: Zhu, Tong
 TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 TITLE OF INVENTION: SAME, AND METHODS OF USE
 FILE REFERENCE: SCRIPI300-3
 CURRENT APPLICATION NUMBER: US/09/938,842A
 CURRENT FILING DATE: 2001-08-24
 PRIOR APPLICATION NUMBER: US 60/227,866
 PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1143
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1143

Alignment Scores:
Pred. No.: 4.94 Length: 1782
Score: 90.00 Matches: 57
Percent Similarity: 30.50% Conservative: 22
Best Local Similarity: 22.01% Mismatches: 66
Query Match: 8.29% Indels: 114
DB: 9 Gaps: 12

US-09-809-545A-2 (1-203) x US-09-938-842A-1143 (1-1782)

QY 5 LyeAlaValAsnProTyrThrAsnGlyTrrPlyLeuAsnProValValGlyAlaValTyr 24
DB 163 CGATCTTTCTCACTTATGCTACCCAGATGATGCTTGATCTTCTGTTTCTAT 216
QY 25 SerProAspPheTyrAla-----GlyThrValLeuLeuCysGln 37
DB 217 GTTCCATATGTGATCAGACGCCCTACTATTATGTTATGGAGTGAACACAGGTTAT 276
QY 38 AlaAsnGlnGlySerSerMetTyrSerGlyPro-----SerSerLeuVal 53
DB 277 ACCAATTCGAGAGTGTATATGATCATCAGAGCTTACGAGAGAGATGATCTTGT 336
QY 54 TyrThrSer-----AlaMetProGlyPheProTyrProAlaThrAlaAlaAla 71
DB 337 TATCCGACAGGATACGGATAGCTGCTCTTCTTACTCAGCAGACAGCCCTGCTCA 396
QY 72 TyrArgGly-----AlaHisLeuArgGlyArgGlyThrValTyr 85
DB 397 CAACCTGGTGGATGATGACAGTGTATGGTGTGACAGATCAGATCCTTCTCTC 456
QY 86 -----Asn 86
DB 457 ACAGCCAGTAGTGGCCTTTCCTGCTGCTGCTGCTTACTCAGAGCAAGCTCTCT 516
QY 87 ThrPheArgAlaAla-----Ala 92
DB 517 ACAACAAAGCAGCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
QY 93 ProProProProLeuProAlaTyrGlyValValTyrGlnGlnProValTyrGlyAsn 112
DB 577 CCGTTAAACCATG-----AATCAGAGTGAAGTGTATGGAAC 615
QY 113 LysLeuLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnPro----- 127
DB 616 AGTGCCTTGGAGTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672
QY 128 -----ThrProAlaThr----- 131
DB 673 GGGTTTACACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 732
QY 132 -----AlaAlaAlaTyrSerAspSerTyrGlyArgVal----- 142
DB 733 GTTTCGGCAGTGGATGATCTATATTAGGCTAAACAATGTAACCTGCGACT 792
QY 142 ----- 142
DB 793 AGGAATCAAACTCTAGCTCAATTCGATTAACAGAGTAGTAGCAGCTGCATCCATG 852
QY 143 -----TyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGly 159
DB 853 ACAGTTATGCACTCAGGGTGTACTACAGACAGATGTCTCCAAACAGTATAGGT 909

Search completed: March 16, 2003, 04:04:03
Job time : 122 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 21:20:45 ; Search time 40 Seconds
(without alignments)
676.247 Million cell updates/sec

Title: US-09-809-545A-2
Perfect score: 1086
Sequence: 1 MTNKKAVNPYTNCKMLNPV.....VLSLSLQSIYOGGYNRPAPY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1086	100.0	203	AAU70146	Rat secreted facto
2	1009.5	93.0	397	AA892573	Human protein sequ
3	1009.5	93.0	440	AAW79112	Human protein SEQ
4	551	50.7	330	AAV81462	Human ataxin-2 bin
5	534	49.2	135	ABG02058	Novel human diagno
6	534	49.2	237	AAW40093	Human polypeptide
7	534	49.2	268	AAW40092	Human polypeptide
8	405.5	37.3	705	ABG29447	Novel human diagno
9	361	33.2	291	AAW41878	Human polypeptide
10	361	33.2	22	AAW41879	Human polypeptide

11	342	31.5	103	22	ABG29445	Novel human diagno
12	288	26.5	366	21	AAV56850	Human RNA binding
13	278	25.6	388	22	ABG26323	Novel human diagno
14	258	23.8	286	22	AAW80096	Human protein SEQ
15	226	20.8	101	22	ABG29446	Novel human diagno
16	178	16.4	151	22	ABG26319	Novel human diagno
17	129	11.9	367	22	ABG29047	Drosophila melanog
18	127.5	11.7	123	22	AAO06790	Human polypeptide
19	116	10.7	318	23	ABW79762	Human Tumour Endot
20	116	10.7	318	22	AAU27790	Human full-length
21	110.5	10.2	219	22	ABW60573	Drosophila melanog
22	110	10.1	474	22	AAU04831	Microspora eve
23	108.5	10.0	244	22	AAW38864	Human polypeptide
24	108.5	10.0	297	22	AAW40650	Human polypeptide
25	103.5	9.5	939	22	AAW78860	Human protein SEQ
26	103.5	9.5	943	22	AAW79844	Human protein SEQ
27	101.5	9.3	950	22	ABW71513	Drosophila melanog
28	100	9.2	830	22	ABW59880	Drosophila melanog
29	98.5	9.1	655	23	ABW57126	Mouse ischaemic co
30	98	9.0	376	22	ABW59035	Drosophila melanog
31	97.5	9.0	828	22	ABW58368	Drosophila melanog
32	97	8.9	179	22	ABW67727	Drosophila melanog
33	96.5	8.9	748	23	ABW92874	Herbicidally activ
34	96	8.8	388	23	ABG53243	Human prostate spe
35	96	8.8	760	18	AAW29450	Programmed cell de
36	96	8.8	760	23	AAU75570	Programmed cell de
37	96	8.8	766	22	ABW64516	Drosophila melanog
38	94	8.7	349	21	AAV5691	Murine Sam68-1 like
39	94	8.7	349	22	AAW49332	Murine SLM-1 prote
40	94	8.7	388	23	ABG61877	Prostate cancer-as
41	94	8.7	576	22	ABW61241	Drosophila melanog
42	93	8.6	444	19	AAW41940	Corn p-hydroxyphen
43	93	8.6	2897	22	ABW58514	Drosophila melanog
44	92.5	8.5	187	22	ABW67725	Drosophila melanog
45	92.5	8.5	362	18	AAW34972	Chimeric Ewing's s

ALIGNMENTS

RESULT 1	
AAU70146	AAU70146 standard; Protein; 203 AA.
XX	XX
AC	AAU70146;
XX	XX
DT	12-MAR-2002 (first entry)
XX	XX
DE	Rat secreted factor protein encoded by DNA clone P0184_D11.
XX	XX
KW	Rat; secreted factor polypeptide; cardiac disease; renal disease; kidney;
KW	inflammatory disease; congestive heart failure; myocarditis; asthma;
KW	dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
KW	myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
KW	atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
KW	renal infarction; hereditary nephritis; polycystic kidney disease;
KW	chronic renal failure; renal vein thrombosis; medullary sponge kidney;
KW	rheumatoid arthritis; osteoarthritis; psoriasis; restenosis;
KW	graft versus host reaction; Crohn's disease; ulcerative colitis;
KW	Alzheimer's disease; gene therapy.
XX	XX
OS	Rattus norvegicus.
PN	WO200174901-A2.
PD	11-OCT-2001.
XX	XX
PF	23-MAR-2001; 2001WO-US09555.
XX	XX
PR	31-MAR-2000; 2000US-193548P.
XX	XX
PA	14-MAR-2001; 2001US-0809545.
XX	XX
XX	(SCIO-) SCIOS INC.

```

XX Stanton LW, White RT;
PI
XX
XX WPI; 2002-010779/01.
DR N-PSDB; AAS94693.
XX
XX Novel secreted factor polypeptide useful for treating cardiac diseases
PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
PT such as asthma, stroke, and rheumatoid arthritis and renal diseases
XX
XX
PS Claim 18; Fig 1; 189pp; English.
XX
XX The invention relates to rat secreted factor polypeptides and the
CC polynucleotides encoding them. The sequences are useful for treating
CC cardiac, renal or inflammatory diseases. These include cardiac diseases
CC such as congestive heart failure, myocarditis, dilated congestive
CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
CC disease, chronic renal failure, renal vein thrombosis and medullary
CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
CC disease. Sequences AAU70146-AAU70178 represent the secreted factor
CC polypeptides of the invention.
XX
SQ Sequence 203 AA;
Query Match 100.0%; Score 1086; DB 23; Length 203;
Query Local Similarity 100.0%; Pred. No. 7.4e-96;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTNKKAVNPYTNCKLNPVGAIVSPDFYAGTVLLCOANQEGSSMYSGPSLVYTSAMPG 60
DB 1 MTNKKAVNPYTNCKLNPVGAIVSPDFYAGTVLLCOANQEGSSMYSGPSLVYTSAMPG 60
QY 61 FFPYATAAAAYRGAHLRGRGRTVYNTFRAAAPPPPIPAYGGVVYQEPYGNKLLQGGYA 120
DB 61 FFPYATAAAAYRGAHLRGRGRTVYNTFRAAAPPPPIPAYGGVVYQEPYGNKLLQGGYA 120
QY 121 ATRYAQPPTATAAAYSDSYGRVYAADPYHHTLAPPTYGVGAMNAFAPLTDKTRSHAD 180
DB 121 ATRYAQPPTATAAAYSDSYGRVYAADPYHHTLAPPTYGVGAMNAFAPLTDKTRSHAD 180
QY 181 VGLVSSLSQASLYGGYNRFAPY 203
DB 181 VGLVSSLSQASLYGGYNRFAPY 203

RESULT 2
AAB92573
ID AAB92573 standard; Protein; 397 AA.
XX
XX AAB92573;
AC
XX
XX 26-JUN-2001 (first entry)
DT
XX
XX Human protein sequence SEQ ID NO:10787.
DE
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
OS
XX
XX EP1074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX
XX 29-JUL-1999; 99JP-0248036.
PR
XX
XX 27-AUG-1999; 99JP-0300253.
PR
XX
XX 11-JAN-2000; 2000JP-0118776.

```

```

PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Claim 8; SEQ ID 10787; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC oligonucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 397 AA;
Query Match 93.0%; Score 1009.5; DB 22; Length 397;
Query Local Similarity 93.6%; Pred. No. 3.6e-88;
Matches 190; Conservative 4; Mismatches 8; Indels 1; Gaps 1;
QY 1 MTNKKAVNPYTNCKLNPVGAIVSPDFYAGTVLLCOANQEGSSMYSGPSLVYTSAMPG 60
DB 196 MTNKKAVNPYTNCKLNPVGAIVSPDFYAGTVLLCOANQEGSSMYSGPSLVYTSAMPG 255
QY 61 FFPYATAAAAYRGAHLRGRGRTVYNTFRAAAPPPPIPAYGGVVYQEPYGNKLLQGGYA 120
DB 256 FFPYATAAAAYRGAHLRGRGRTVYNTFRAAAPPPPIPAYGGVVYQDGRYAD-YYGGYA 314
QY 121 ATRYAQPPTATAAAYSDSYGRVYAADPYHHTLAPPTYGVGAMNAFAPLTDKTRSHAD 180
DB 315 ATRYAQPPTATAAAYSDSYGRVYAADPYHHTLAPPTYGVGAMNAFAPLTDKTRSHAD 374
QY 181 VGLVSSLSQASLYGGYNRFAPY 203
DB 375 VGLVSSLSQASLYGGYNRFAPY 397

RESULT 3
AAM79112
ID AAM79112 standard; Protein; 440 AA.
XX
XX AAM79112;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human protein sequence SEQ ID NO 1774.
DE
XX
XX

```


CC (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX
SQ Sequence 330 AA;

Query Match 50.7%; Score 551; DB 21; Length 330;
Best Local Similarity 97.1%; Pred. No. 2e-44; 2; Indels 0; Gaps 0;
Matches 102; Conservative 1; Mismatches 2;

Oy 1 MTNKAIVNVTGKMLNPVGAIVSPDPFAGTVLLCOANQEGSSMYSGPSSLVYTSAMP 60
Db 196 MTNKTIVNVTGKMLNPVGAIVSPDPFAGTVLLCOANQEGSSMYSGPSSLVYTSAMP 255

Oy 61 FFPYATAAAAYRGALHRCGRVTYNTFPAAPPPPIPAVGAVY 105
Db 256 FFPYATAAAAYRGALHRCGRVTYNTFPAAPPPPIPAVGAVY 300

RESULT 5
ABG02058
ID ABG02058 standard; Protein; 135 AA.

AC ABG02058;
DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #2049.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS66245.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

Claim 20; SEQ ID NO 32417; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 135 AA;

Query Match 49.2%; Score 534; DB 22; Length 135;
Best Local Similarity 64.0%; Pred. No. 2.7e-43;
Matches 105; Conservative 6; Mismatches 23; Indels 30; Gaps 4;

Oy 1 MTNKAIVNVTGKMLNPVGAIVSPDPFAGTVLLCOANQEGSSMYSGPSSLVYTSAMP 60
Db 1 MTNKTIVNVTGKMLNPVGAIVSPDPFAGTVLLCOANQEGSSMYSGPSSLVYTSAMP 255

Oy 61 FFPYATAAAAYRGALHRCGRVTYNTFPAAPPPPIPAVGAVYQEPYGNKLLGGYA 120
Db 34 FFPYPTGTAVAYRGALHRCGRVTYNTFPAAPPPPIPAVGAVYQDPGYAE-IYGYA 92

Oy 121 AYRYAOPPTATTAAYSDSYGRYV-AADPYHHTLAPPTYGVAM 163
Db 93 AYRYAOPFA-AAYSDSYGRVYAAADPYHHTIGPATYSIGTM 135

RESULT 6
AAM40093
ID AAM40093 standard; Protein; 237 AA.

AC AAM40093;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3238.

KW Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

XX 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Auand V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

DR WPI: 2001-442253/47.

DR N-PSDB; AAS59249.

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 5; SEQ ID NO 3238; 10078pp; English.

CC The invention relates to human nucleic acids (AAS57798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nocotropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

CC Sequence 237 AA;

Query Match 49.2%; Score 534; DB 22; Length 237;

Best Local Similarity 64.0%; Pred. No. 5.5e-43; Mismatches 23; Indels 30; Gaps 4;

Matches 105; Conservative 6; Mismatches 23; Indels 30; Gaps 4;

CC 1 MTNKKAVNPYTNMGKLNPPVGAIVSPDFYAGTVLLCOANQCGSSWYSGPSSLVYTSAMPG 60

CC 103 MTNKKGNPPTNGKMLNPVGAIVGPEFYAVT-----G 135

CC 61 FPPYPAATAAAYRGAGHLRGRGRTVYNTFRRAAPPPIPAYGVVYQBPVGNKLLQGGYA 120

CC 136 FPPYTTGTAAVYRGAGHLRGRRAVYNTFRRAAPPPIPTYGAVVYQDGFYGA- IYGGYA 194

CC 121 AYRYAOPTPATAAAYSDSYGRVY-AADPYHHTLAPATYGVGAM 163

CC 195 AYRYAOPA-AAAAAYSDSYGRVYAAADPYHHTIGPATYSIGTM 237

CC Db

CC RESULT 7

CC ID AAM40092

CC AAM40092 standard; Protein: 268 AA.

CC XX AAM40092;

CC XX 22-OCT-2001 (first entry)

CC Human polypeptide SEQ ID NO 3237.

CC Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

CC peripheral nervous system; neuropathy; central nervous system; CNS;

CC Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

CC amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

CC chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

CC leukaemia.

CC XX Homo sapiens.

CC OS WO200153312-A1.

CC XX 26-JUL-2001.

CC PF 26-DEC-2000; 2000MO-US34263.

CC XX 21-JAN-2000; 2000US-0488725.

CC PR 25-APR-2000; 2000US-0582317.

CC PR 09-JUL-2000; 2000US-0598042.

CC PR 19-JUL-2000; 2000US-0620312.

CC PR 03-AUG-2000; 2000US-0653450.

CC PR 14-SEP-2000; 2000US-0662191.

CC PR 19-OCT-2000; 2000US-0693036.

CC PR 29-NOV-2000; 2000US-0727344.

CC XX (HYSE-) HYSEQ INC.

CC PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,

CC PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J,

CC PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR N-PSDB; AAI59248.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Example 5; SEQ ID NO 3237; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic;

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.

CC Sequence 268 AA;

CC XX

CC Query Match 49.2%; Score 534; DB 22; Length 268;

Best Local Similarity 64.0%; Pred. No. 6.5e-43; Mismatches 23; Indels 30; Gaps 4;

Matches 105; Conservative 6; Mismatches 23; Indels 30; Gaps 4;

CC 1 MTNKKAVNPYTNMGKLNPPVGAIVSPDFYAGTVLLCOANQCGSSWYSGPSSLVYTSAMPG 60

CC 134 MTNKKGNPPTNGKMLNPVGAIVGPEFYAVT-----G 166

CC 61 FPPYPAATAAAYRGAGHLRGRGRTVYNTFRRAAPPPIPAYGVVYQBPVGNKLLQGGYA 120

CC 167 FPPYTTGTAAVYRGAGHLRGRRAVYNTFRRAAPPPIPTYGAVVYQDGFYGA- IYGGYA 225

CC 121 AYRYAOPTPATAAAYSDSYGRVY-AADPYHHTLAPATYGVGAM 163

CC 226 AYRYAOPA-AAAAAYSDSYGRVYAAADPYHHTIGPATYSIGTM 268

CC Db

CC RESULT 8

CC ID ABG29447

CC ABG29447 standard; Protein: 705 AA.

CC XX ABG29447;

CC XX 18-FEB-2002 (first entry)

CC Novel human diagnostic protein #29438.

CC Human; chromosome mapping; gene mapping; gene therapy; forensic;

CC food supplement; medical imaging; diagnostic; genetic disorder.

CC XX Homo sapiens.

CC OS WO200175067-A2.

CC XX 11-OCT-2001.

CC PD 30-MAR-2001; 2001MO-US08631.

CC PF 31-MAR-2000; 2000US-0540217.

CC PR 23-AUG-2000; 2000US-0649167.

CC XX (HYSE-) HYSEQ INC.

CC PA Drmanac RT, Liu C, Tang YT;

CC XX

DR WPI: 2001-639362/73.
DR N-PSDB: AAS93634.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID NO 59606; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostic as expressed sequence tags
XX for identifying expressed genes. (II) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 705 AA;

Query Match 37.3%; Score 405.5; DB 22; Length 705;

Best Local Similarity 34.4%; Pred. No. 4,4e-30; Indels 181; Gaps 7;

Matches 105; Conservative 9; Mismatches 10; Indels 181; Gaps 7;

QY 53 VTSMPGPPYPAATTAAYRGALRGRTVYNTFRRAAPPPIPAYGV----- 103
DB 130 VEINSPGPPYPAATTAAYRGALRGRTVYNTFRRAAPPPIPAYGVLMNVLDPY 249
QY 104 -----VY--OE-----PYVG 111
DB 250 FALSDDDFNILLVCIRVYTGQESTLCHCPAQCKRDQLKMFVDAQCPQLGSPV-- 307
QY 112 NKLLOGGVAAYVAOPTPAATAAYSP----- 137
DB 308 PRLGGGVAAYVAOPTPAATAAYSDRLSLMGDSSEIVKGRNQFVVAADDEISCNTSA 367
QY 138 ----- 137
DB 366 GCIYPSYQLEASSGVGCVTSAPHSIFPTSQPHAAASYIKTLDSGDAMDKIDTPAPASMEF 427
QY 138 -----SYGRVYAADPYHHTLAPAPTYGVGAMNAFPLTDAKTRSHADVGLVSLQAS 191
DB 428 SFOEGSSYGRVYAADPYHALAPAPTYGVGAMN-----RQY-----SSLCVS 469
QY 192 IYQGG 196
DB 470 VCTGG 474

RESULT 9

AA041878
ID AA041878 standard; Protein; 291 AA.

XX AA041878;

XX 22-OCT--2001 (first entry)

XX Human polypeptide SEQ ID NO 6809.

KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0633450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HSE-) HYSEQ INC.

XX Tang YT, Liu C, Saundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;
XX Zhao QH, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB; AAI61034.

XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -

XX Example 2; SEQ ID NO 6809; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AA036642-AA042213) with nootropic,
XX immunosuppressant and cyostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.

SO Sequence 291 AA;

Query Match 33.2%; Score 361; DB 22; Length 291;

Best Local Similarity 55.5%; Pred. No. 2.5e-26; Indels 34; Gaps 2;

Matches 71; Conservative 5; Mismatches 18; Indels 34; Gaps 2;

QY 1 WTKKAVNDYNTGKLNPNVGAIVSPDYAGTVLLCQANQEGSSMTSGSLVYSAMPG 60

DB 147 MTNKKTGNYNTGKLNPNVGAIVSPDYAGTVLLCQANQEGSSMTSGSLVYSAMPG 179

QY 61 FPPYPAATTAAYRGALRGRTVYNTFRRAAPPPIPAYGVYQEPYGNKLLGGYA 120

DB 180 FPPYPTTGAVAVRGALRGRTVYNTFRRAAPPPIPAYGVYQEPYGNKLLGGYA 236

QY 121 AYRYAOPT 128

DB 237 ----TQPT 240

RESULT 10
AAM41879
ID AAM41879 standard; Protein; 291 AA.
AC AAM41879;
XX
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 6810.
XX
XX Human, noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX Homo sapiens.
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX DR N-PSDB; AAI61035.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6810; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA15798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with noctropic;
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: immune system suppression;
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX
XX Sequence 291 AA;
XX
XX Query Match 33.2%; Score 361; DB 22; Length 291;
XX Best Local Similarity 55.5%; Pred. No. 2.5e-26;
XX Matches 71; Conservative 5; Mismatches 18; Indels 34; Gaps 2;
XX
XX 1 MTNKAANPNTNGWKLNPVGAIVSPDFYAGTVLLCQANQEGSSMYSGPSSLIIVTSAMP 60

Db 147 MTNKTGNPNTNGWKLNPVGAIVSGPPEYAVT-----G 179
Oy 61 FPPYPAATAAAYRGAAHJRGRTVYNTFRAAPPPIPAYGVVYQEPYGNKLLGGYA 120
Db 180 FPPYTTGTAIVYRGAAHJRGRTAVYNTFRAAPPPIPTYGAIVYQDGFYGAETLEA--- 236
Oy 121 AYRYAOPR 128
Db 237 ----TQPT 240
RESULT 11
ABG29445
ID ABG29445 standard; Protein; 103 AA.
XX
XX ABG29445;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #29436.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX OS
XX WO200175067-A2.
XX PN
XX 11-OCT-2001.
XX PD
XX 30-MAR-2001; 2001WO-US08631.
XX PF
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS93632.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX Claim 20; SEQ ID NO 59804; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 103 AA;

Query Match 31.5%; Score 342; DB 22; Length 103;
 Best Local Similarity 97.0%; Pred. No. 4.3e-25;
 Matches 64; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 138 SYGRVYADPHHTHTAPPTVGVGMNAPFLTDKTRSHADVDGLVLSLQASITRGY 197
 DB 4 SYGRVYADPHHTHTAPPTVGVGMNAPFLTDKTRSHADVDGLVLSLQASITRGY 63

OY 198 NRPAPY 203
 DB 64 NRPAPY 69

RESULT 12
 AAY56850
 ID AAY56850 standard; Protein; 366 AA.
 XX
 AC AAY56850;
 XX
 DT 11-APR-2000 (first entry)
 XX
 DE Human RNA binding protein (RNABP)-2 (clone 1250374).
 XX
 KM RNA binding protein; RNABP; cancer; immune disorder; AIDS; human;
 KM developmental disorder; acquired immunodeficiency syndrome; RNABP-2;
 KM inflammation; allergy; diabetes mellitus; seizure disorder.
 XX
 OS Homo sapiens.
 XX
 PN US6020164-A.
 XX
 PD 01-FEB-2000.
 XX
 PE 21-OCT-1998; 98US-0176657.
 XX
 PR 21-OCT-1998; 98US-0176657.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Bandman O, Corley NC, Lu DM, Baughn MR, Tang YT, Guegler KJ;
 DR WP1; 2000-146885/13.
 DR N-PSDB; AA246827.
 XX
 PT Isolated and purified polynucleotide for modulating the expression of
 PT human RNA binding proteins which play a role in cancer; immune
 PT disorders and developmental disorders -
 XX
 PS Examples; Column 45-48; 39pp; English.
 XX
 CC The present invention provides human RNA binding proteins (RNABP) and
 CC polynucleotides encoding the proteins. The polypeptides are useful for
 CC modulating the expression of human RNA binding proteins (RNABP) which
 CC play a role in cancer, immune disorders and developmental disorders.
 CC Disorders associated with a decrease of RNABP include: cancers such as
 CC adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,
 CC teratocarcinoma, and, in particular, cancers of the adrenal gland,
 CC bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,
 CC gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,
 CC pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,
 CC testis, thymus, thyroid, and uterus, immune disorders such as acquired
 CC immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory
 CC distress syndrome, allergies, ankylosing spondylitis, amyloidosis,
 CC anemia, achma, atherosclerosis, autoimmune hemolytic anemia, autoimmune
 CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's
 CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,
 CC emphysema, episodic lymphopenia with lymphocytocytosis, erythroblastosis
 CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,
 CC Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,
 CC hyperostosis, hyperostosis, hyperostosis, hyperostosis, hyperostosis,
 CC myasthenia gravis, myocardial or pericardial inflammation,
 CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,

CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's
 CC syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic
 CC sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner
 CC syndrome, complications of cancer, hemodialysis, and extracorporeal
 CC circulation, viral, bacterial, fungal, parasitic, protozoal, and
 CC helminthic infections, trauma, X-linked agammaglobulinemia of Bruton,
 CC common variable immunodeficiency (CVID), Diego's syndrome (thymic
 CC hypoplasia), thymic dysplasia, isolated immunoglobulin (Ig) A deficiency,
 CC severe combined immunodeficiency disease (SCID), immunodeficiency with
 CC thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi
 CC syndrome, chronic granulomatous diseases, hereditary angioneurotic edema,
 CC and immunodeficiency associated with Cushing's disease; and developmental
 CC disorders such as renal tubular acidosis, anemia, Cushing's syndrome,
 CC achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,
 CC epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia,
 CC genitourinary abnormalities, and mental retardation), Smith-Magenis
 CC syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis,
 CC hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-
 CC Tooth disease and neurofibromatosis, hypochromidism, hydrocephalus,
 CC seizure disorders such as Sydenham's chorea and cerebral palsy, spina
 CC bifida, anencephaly, craniofacial dysmorphism, congenital glaucoma, cataract,
 CC sensorineural hearing loss, and any disorder associated with cell
 CC growth and differentiation, embryogenesis, and morphogenesis involving
 CC any tissue, organ, or system of a subject, e.g., the brain, adrenal
 CC gland, kidney, skeletal or reproductive system. The present sequence
 CC represents the RNABP-2 polypeptide.
 XX
 SQ Sequence 366 AA;

Query Match 26.5%; Score 288; DB 21; Length 366;
 Best Local Similarity 48.9%; Pred. No. 3.2e-19;
 Matches 68; Conservative 11; Mismatches 54; Indels 6; Gaps 4;

OY 1 MTNKAANPYTNGWMLANVGVAVSPDFAGTLLCOAN--QEGSSMYSGPSSL--VYTS 56
 DB 190 MTNKAANPYTNGWMLANVGVAVSPDFAGTLLCOAN--QEGSSMYSGPSSL--VYTS 249

OY 57 AMPGPPYP-AATTAAYAGAHILRGRTVYNTFRAPAAPPPYPAYGVYQEPVGNKLL 115
 DB 250 IIPGPPYPTAATTAAYAGAHILRGRTVYNTFRAPAAPPPYPAYGVYQEPVGNKLL 308

OY 116 OGGYAAHYAOPTPATMAA 134
 DB 309 OGPPLLOPLPLTYTMA 327

RESULT 13
 ABG26323
 ID ABG26323 standard; Protein; 388 AA.
 XX
 AC ABG26323;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #26314.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PE 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.
DR N-PSDB; AAS90510.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 56682; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 388 AA;
Query Match 25.6%; Score 278; DB 22; Length 388;
Best Local Similarity 55.6%; Pred. No. 3.2e-18;
Matches 60; Conservative 9; Mismatches 33; Indels 6; Gaps 4;
QY 1 MTNKKAVNPYTNMGKLNPNVGAVYSPDFYAGTVLLCQAN--QEGSSMYSGPSS--VYTS 56
DB 186 MTNKKMVPYVNGWKLSPVAVGVPBELVLAASSFQADVSLGNDAAVPLSGRGINTYIPL 245
QY 57 AMPGFYVP-AATAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGV 103
DB 246 IIPGFYTAATTAARFAGHLRGRGRVYGAVR-AVPEPTAIAPGV 292
RESULT 14
AAM80096
ID AAM80096 standard; Protein; 286 AA.
XX
XX AAM80096;
XX
XX 06-NOV-2001 (first entry)
XX
XX
DE Human protein SEQ ID NO 3742.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.
XX

PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Weinman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX N-PSDB; AAK53229.
DR
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
PT
XX
XX Claim 20; Page 423; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
SQ Sequence 286 AA;
Query Match 23.8%; Score 258; DB 22; Length 286;
Best Local Similarity 92.2%; Pred. No. 1.7e-16;
Matches 47; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MTNKKAVNPYTNMGKLNPNVGAVYSPDFYAGTVLLCQANQEGSSMYSGPSS 51
DB 219 MTNKKTVNPYTNMGKLNPNVGAVYSPDFYAGTVLLCQANQEGSSMYSPST 269
RESULT 15
ABG29446
ID ABG29446 standard; Protein; 101 AA.
XX
XX ABG29446;
XX
XX 18-FEB-2002 (first entry)
XX
XX
DE Novel human diagnostic protein #29437.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 N-PSDB; AAS93633.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20; SEQ ID No 59805; 103pp; English.

XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 101 AA;

Query Match 20.8%; Score 226; DB 22; Length 101;
 Best Local Similarity 50.0%; Pred. No. 5, 2e-14;
 Matches 52; Conservative 6; Mismatches 20; Indels 26; Gaps 3;

OY 69 AAAYRGALRGKRTVYNTFRRAAPPPIPAYGVVYQBPVGNKLLGGYAAVRYAQP 128
 DB 1 AAAYRGALRGKRTVYNTFRRAAPPPIPAYGVVYQDPFYG----- 43
 OY 129 PATAAVSDSYG-RVYAADPYHHTLAPPTYGV-GAMNAFAPLT 170
 DB 44 -----ADLYGVMLHTATPSLPPLPLPTVTVTDEFMTFSPIS 80

Search completed: March 12, 2003, 22:24:15
 Job time : 40 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:23:36 ; Search time 25 Seconds
(without alignments)
780.612 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MTNKKAVNPYTNQWKLNPV.....VLSSLOASTYOGYNRPAPY 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR_73:*

1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	9.5	338	2	C75607 conserved hypothet
2	103	9.5	381	2	KM8 protein rat
3	103	9.5	3124	1	collagen alpha 1(X
4	99	9.1	542	2	hypothetical prote
5	98.5	9.1	655	1	RNA-binding protei
6	98	9.0	316	2	hypothetical prote
7	98	9.0	376	2	segmentation prote
8	97	8.9	811	2	probable transmem
9	96.5	8.9	139	2	extensin precursor
10	96.5	8.9	748	2	hypothetical prote
11	96	8.8	760	2	eye cell developme
12	95	8.7	418	2	hypothetical prote
13	94.5	8.7	469	2	hypothetical prote
14	94	8.7	219	2	adhesin protein -
15	94	8.7	967	2	T20H2.10 protein
16	93.5	8.6	320	2	chorion protein s3
17	93	8.6	387	2	transcription fact
18	93	8.6	495	1	probable oxidoredu
19	92.5	8.5	348	2	RNA-binding protei
20	92.5	8.5	656	1	hypothetical prote
21	91.5	8.4	325	2	probable PPE prote
22	91.5	8.4	409	2	E1B-55KDa-associat
23	91.5	8.4	856	2	zinc finger RNA bi
24	91.5	8.4	1052	2	eyelid - fruit fly
25	91.5	8.4	2715	2	hypothetical prote
26	91	8.4	164	2	probable secreted
27	91	8.4	748	2	hypothetical prote
28	90.5	8.3	502	2	branched-chain ami
29	89	8.2	318	2	T00626

30	89	8.2	463	2	C70931 probable PPE prote
31	89	8.2	805	2	T25795 hypothetical prote
32	88.5	8.1	390	2	urea/short-chain a
33	88	8.1	2639	2	fibroin - Chinese
34	87.5	8.1	912	2	D72644 hypothetical prote
35	87.5	8.1	1685	2	T43217 RNA polymerase (EC
36	87	8.0	465	2	G02738 FREAC-4 - human
37	87	8.0	482	2	C84330 pantothenate perme
38	86.5	8.0	262	2	A54889 IGE-binding protei
39	86.5	8.0	4302	2	A38971 polycystic kidney
40	86	7.9	568	2	D75627 potassium-transport
41	86	7.9	860	1	EAMS elastin precursor
42	85.5	7.9	345	2	A48462 dense granule prot
43	85.5	7.9	1279	2	T13613 hypothetical prote
44	85	7.8	268	2	A49303 homeotic protein C
45	85	7.8	382	1	SUBSN subtilisin (EC 3.4

ALIGNMENTS

RESULT 1
C75607 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:/Species: Deinococcus radiodurans
C:/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:/Accession: C75607
R:/White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S. J. Smith, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Moh. Science 286, 1571-1577, 1999
A:/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:/Reference number: A75250; MUID:20036896; PMID:10567266
A:/Accession: C75607
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-338 <WHI>
A:/Cross-references: GB:AB001862; GB:AE001825; NID:G6460468; PIDN:APF12236.1; PID:G6460530
A:/Experimental source: strain R1
C:/Genetics:
A:/Gene: DR0120
A:/Map position: 2
C:/Superfamily: ymbw protein

Query Match 9.5%; Score 103; DB 2; Length 338;
Best Local Similarity 24.9%; Pred. No. 0.3; 78; Indels 36; Gaps 8;
Matches 45; Conservative 22; Mismatches 1; A:SVFDGLAE 161

QY 20 VGAVYSPDFVAGTVLLCOANQEGSSMTSGPSLVYTSAMPG-FPYPAATMAAYRGANLR 78
DB 105 LGIGRAPGPTDRTALALRSRELAADDFPTQIALLRFAQETWMP--ACSVFDGLAE 161

QY 79 GGRGTVTNTPFAAPPPPIPAVGGVYIOEPYGNKLOGGTA-AVRYAQTPATPAAYSD 137
DB 162 PRGVYL-----PRLMILGSSLYGAQLAGE--LGVGAFAFYHFSQEDPAPV----- 203

QY 138 SYGRVYADPPYHHTLAPR---TYGVGMNAFAPLTDKTSHADDDVGLVSSLOASTYQ 194
DB 204 -----AVQTYHRRHFRPGLAEPYALIGVNLA-----APTDEARDLSLTSAAISGLTLS 253

QY 195 G 195
DB 254 G 254

RESULT 2
JC4647
KM8 protein - rat
N/Alternate names: basic helix-loop-helix protein homolog
C:/Species: Rattus norvegicus (Norway rat)
C:/Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C:/Accession: JC4647
R:/Kume, H.; Maruyama, K.; Tomita, T.; Iwatsubo, T.; Saido, T.C.; Ohtsuka, K.

Biochem. Biophys. Res. Commun. 219, 526-530, 1996
A>Title: Molecular cloning of a novel basic helix-loop-helix protein from the rat brain.
A'Reference number: J04647; MUID:96193685; PMID:8605021
A'Accession: J04647
A'Molecule type: DNA
A'Residuals: 1-381 <KUM>
A'Cross-references: DDBJ:D82668; NID:g1166397; PIDN:BA11615.1; PID:g1166398
A'Experimental source: brain
C'Comment: This protein is involved in synaptic plasticity, and has a role specific to r-loop-helix domain.
C'Keywords: brain
F:122-134/Region: basic
Query Match 9.5%; Score 103; DB 2; Length 381;
Best Local Similarity 22.8%; Pred. No. 0.34;
Matches 39; Conservative 24; Mismatches 76; Indels 32; Gaps 4;
Qy 14 WKLNPFVGVAVSPDFYAGTVLCOA-----NQESSMYSQPSL 52
Db 171 WALSELRSGRDLDVSYVQTLCKLSQPTNLVAGCLQNSRNFLEQGDAGDAAFTRAR 230
Qy 53 VYTSAPRGFPYPA-----TAAAYRGALRGGR-TVNT-FRAAAPPPPAYG 101
Db 231 VARSAMHPRPFCSLAGSARPPAAGRRGALRTHTGYCAVETLYAAGGGAAPDVN 290
Qy 102 GVYQBPVYGNKLLQGYAARYAOPTRPATAAYSDSYGRVYADPYHNTL 152
Db 291 SSEYEGPLSPRLCLNGNFSIAKQSSPDHEKSYHSMYSLPGRAPGHCL 341
RESULT 3
A00020
collagen alpha 1(XII) chain precursor - chicken
N:Alternate names: fibrochimerin
C:Species: Gallus gallus (Chicken)
C>Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #ext. change 19-Jan-2001
C'Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shimomura, T.; Tanaka, H.; Nishida, Y.; Obari, J. Cell Biol. 115, 209-221, 1991
A>Title: The complete primary structure of type XII collagen shows a chimeric molecule with house region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.
A'Reference number: A40020; MUID:92011862; PMID:1518137
A'Accession: A40020
A'Molecule type: mRNA
A'Residuals: 1-3124 <YAM>
A'Cross-references: GB:D00824; NID:9222810; PIDN:BA00701.1; PID:9222811
A>Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and, R:Gordon, M.K.; Gerecke, D.R.; Dudley, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989
A>Title: Type XII collagen. A large multidomain molecule with partial homology to type I A'Reference number: A34485; MUID:90062079; PMID:2584192
A'Accession: A34485
A'Molecule type: mRNA
A'Residuals: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A'Cross-references: EMBL:005137; NID:g211284; PIDN:AAA4635.1; PID:g211285
A'Accession: B34485
A'Molecule type: protein
A'Residuals: 2772-2792, 2846-2873 <GOR2>
R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A>Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA A'Reference number: A28037; MUID:87317590; PMID:3476925
A'Accession: A28037
A'Molecule type: mRNA
A'Residuals: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
A'Cross-references: EMBL:M17315; NID:g211649; PIDN:AAA48718.1; PID:g211650
A>Note: this sequence has been revised in reference A34485
R:Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A>Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of A'Reference number: S23814; MUID:92362621; PMID:1323460
A'Accession: S23814
A'Molecule type: protein

A'Residuals: 'X', 1333, 'Q', 1335-1347, 1914-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517 <R'Dublet, B.; van der Rest, M.
J. Biol. Chem. 262, 17724-17727, 1987
A>Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-A'Reference number: S22254; MUID:88087065; PMID:3121603
A'Accession: S22254
A'Molecule type: protein
A'Residuals: 2831-2832, 'T', 2834, 'R', 2836-2843, 3002-3014 <DUB>
R:Trub, J.; Trub, B.
Biochem. Biophys. Acta 1171, 97-98, 1992
A>Title: The two splice variants of collagen XII share a common 5' end.
A'Reference number: S28811; MUID:93042014; PMID:1420368
A'Accession: S28811
A'Molecule type: mRNA
A'Cross-references: EMBL:X67327
C'Genetics: 2845/3; 2863/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
A'Intons: 2845/3; 2863/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
C'Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von Willebrand factor; cell binding; cell binding; coiled coil; connective tissue; disulfide
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F:24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted <F24-114/Domain: IIA #status predicted <IIA>
F:24-105/Domain: fibronectin type III repeat homology <FN3A>
F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>
F:332-425/Domain: IIB #status predicted <IIB>
F:332-414/Domain: fibronectin type III repeat homology <FN3B>
F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
F:629-1178/Domain: IIC #status predicted <IIC>
F:630-711/Domain: fibronectin type III repeat homology <FN3C>
F:721-802/Domain: fibronectin type III repeat homology <FN3D>
F:812-895/Domain: fibronectin type III repeat homology <FN3E>
F:905-986/Domain: fibronectin type III repeat homology <FN3F>
F:1095-1076/Domain: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
F:1197-1261/Domain: von Willebrand factor type A repeat homology <VWA3>
F:1384-2295/Domain: IID #status predicted <IID>
F:1384-1465/Domain: fibronectin type III repeat homology <FN3I>
F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>
F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>
F:1655-1738/Domain: fibronectin type III repeat homology <FN3L>
F:1756-1838/Domain: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>
F:2118-2199/Domain: fibronectin type III repeat homology <FN3Q>
F:2135-2490/Domain: von Willebrand factor type A repeat homology <VWA4>
F:2325-2440/Region: cell adhesion #status predicted
F:2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted
F:2751-2802/Domain: collagenous COL2 #status predicted <COL2>
F:2899-2901/Region: cell attachment (R-G-D) motif
F:2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>
F:2946-3048/Domain: collagenous COL1 #status predicted <COL1>
F:3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>
F:32, 1006, 1032, 1044, 1512, 1767, 2210, 2273, 2532, 2683/Binding site: carboxylate (Asn) (Cova F:2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #status predicted
Query Match 9.5%; Score 103; DB 1; Length 3124;
Best Local Similarity 25.1%; Pred. No. 3.8;
Matches 42; Conservative 15; Mismatches 46; Indels 64; Gaps 8;
Qy 62 PYPAATAAAAYR---GAHLRGGRV-----VTPFAAAPPPIPAVGV 104
Db 2094 PV-KITVAIVYDGGGGLTGNGRTVGLIPONIIYITDEWYFRVSWDPSFVLGYKI 2152
Qy 105 YQBPVYGNKLLQGYAARYAOPTRPATAAYSDSYGRVYADPYHNTLAPPTVGV 161
Db 2153 YKPVQSNPMEVPGVEVSYTL-----HNLSPSTTYDV- 2186
Qy 162 ANNAFA-----PLTDAKTRSHADVGLVLSLQASTYGGYNRF 200

A:Reference number: A26636; MUID:87218536; PMID:2884106
 A:Accession: A26636
 A:Molecule type: mRNA
 A:Residues: 1-299 'L', 301-376 <FRA>
 A:Cross-references: GB:X05138; NID:g7957; PIDN:CAA28784.1; PID:g7958
 C:Genetic8:
 A:Gene: eve (even-skipped)
 A:Cross-references: FlyBase:FBgn0000606
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:71-127/Domain: homeobox homology <HOK>

Query Match 9.0%; Score 98; DB 2; Length 376;
 Best Local Similarity 24.3%; Pred. No. 0.88;
 Matches 45; Conservative 8; Mismatches 56; Indels 76; Gaps 7;

QY 22 AYSPDFAGTLLCOANQEGSS--MYSGPSLVYTSAMPGPYPATNAAYRGALRG 80
 DB 138 AYSDPAFASTLQAAANSVG-----MYPYAPALAAANAAAA----- 177
 QY 81 GRTVYTFPAAPPP-----PIPAYGVVYQEPVYGNKLLQGGVAAVRYAQTPTAT 131
 DB 178 --VANTPMWATGMPMGMPQMPTMGPHSGHGHSPYQG-----YKXT----- 220
 QY 132 AAAYSDSYGRVYAADPYHTLAPAPTYGVGAMNAPLTDAKTRSHADVGLVSLQAS 191
 DB 221 -----FYHHPARAPAPHPAG-----PHHHPHMGSSATGS 251
 QY 192 IYQGG 196
 DB 252 SYSAG 256

RESULT 8

T36581
 Probable transmembrane protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36581
 R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: 221575
 A:Accession: T36581
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-811 <OLIT>
 A:Cross-references: EMBL:AL049826; PIDN:CAB42720.1; GSPDB:GM00070; SCODEB:SCH24.16C
 A:Experimental source: strain AJ12
 C:Genetic8:
 A:Gene: SCODEB:SCH24.16C

Query Match 8.9%; Score 97; DB 2; Length 811;
 Best Local Similarity 27.5%; Pred. No. 2.6;
 Matches 49; Conservative 14; Mismatches 63; Indels 52; Gaps 11;

QY 24 YSPDFAGTLLCOANQEGSS--MYSGPSLVYTSAMPGPYPATNAAYRGALRG 81
 DB 140 HRPDAFAH---LFRDQGGGHSYDDQAAPAPAPG-PYGGAAAPGQY----- 186
 QY 82 KTVYTFPAAPPPPIPAYGVVYQEPVYGNKLLQGGVAAVRYAQTPTAAAYSISYGR 141
 DB 187 -----GAPPPPPGQYGGHGGPDP-YGNAQTHGQYGG-----SAQYGS--- 224
 QY 142 VYAADPYHTLAPAP-TYGVGAM--NAPAPLTDAKTRSHADVGLVSLQASISYQ 196
 DB 225 --AATFGQA-APAGQIVAPTPPEAAEAPLQEPPEPPAS-----QAAPKKG 270

RESULT 9

S61885
 extensin precursor (clone 6Pext1.2) - wood tobacco
 C:Species: Nicotiana glauca (wood tobacco)
 C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 19-Jul-2002

C:Accession: S61885
 R:Parmentier, Y.; Durr, A.; Marbach, J.; Hirsinger, C.; Criqui, M.C.; Fleck, J.; Jamet, Plant Mol. Biol. 29, 279-292, 1995
 A:Title: A novel wound-inducible extensin gene is expressed early in newly isolated prot.
 A:Reference number: S61885; MUID:96046747; PMID:7579179
 A:Accession: S61885
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-139 <PAR>
 A:Cross-references: EMBL:X70343
 C:Superfamily: proline-rich protein 3

Query Match 8.9%; Score 96.5; DB 2; Length 139;
 Best Local Similarity 26.1%; Pred. No. 0.37;
 Matches 37; Conservative 12; Mismatches 52; Indels 41; Gaps 6;

QY 34 ILCOANQEGSS-----MYSGPSLVYTSAMPGPYPATNAAYRGALRGRTY 85
 DB 22 LECKANYYSSPPPTKYYSSPPPYKYSPPPLP-----IY 62
 QY 86 NT-----FRAAPPPPIPAYGVVYQEPVYGNKLLQGGVAAVRYAQTPTAAAYSISY 140
 DB 63 RSPPPVYKYSPPPIYKYSPPPPPVYKSP---PPVYKYSPPPPPVYKSP 117
 QY 141 RYVYADP---YHHTLAPAPTY 158
 DB 118 PVKSPPPPYHYTSPPPHY 139

RESULT 10

T04011
 hypothetical protein TSL19.200 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 C:Accession: T04011
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Meyer, K.F.X
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15184
 A:Accession: T04011
 A:Molecule type: DNA
 A:Residues: 1-748 <BEV>
 A:Cross-references: EMBL:AL049481
 A:Experimental source: cultivar Columbia; BAC clone TSL19
 C:Genetic8:
 A:Map position: 4
 A:Introns: 67/2; 209/3; 271/3; 305/3; 329/3; 367/3; 691/1
 A>Note: TSL19.200

Query Match 8.9%; Score 96.5; DB 2; Length 748;
 Best Local Similarity 26.9%; Pred. No. 2.6;
 Matches 46; Conservative 16; Mismatches 70; Indels 39; Gaps 7;

QY 18 PVVGVYSPDFAGTLLCOANQEGSSMYSGP-----SLVYTSAMPGPYPATNAAYR 73
 DB 584 PYQATPPAIPSTGSTMARPQQQYGYTSDDGVQQQTYPSISAPPSDAYNNGTOTPATG 643
 QY 74 GAHLRGRTVYNTF-RAAAPPPPIPAYGVV-----YQEPVYGNKLLQGGVAYR 123
 DB 644 PAYQOQSVQPASTYDQGAQAAAGYGQVAPRGTYTPRSQAYSGQ-----AAYS 697
 QY 124 VYQPT-----PATPAAYSDSYGRVYAADPYHTLAPAPTYGVGAMNAPFA 167
 DB 698 QAAPQTQYEQQPATQAA-----VYATAP-----GTAPVYQSPQSAFYA 736

RESULT 11

A45174
 eye cell development gene eye protein, splice form 1 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
 C:Accession: A45174
 R:Bonini, N.M.; Leiserson, W.M.; Benzer, S.
 Cell 72, 379-395, 1993

A>Title: The eyes absent gene: genetic control of cell survival and differentiation in *C. elegans*
A:Reference number: A45174; MUID:93161413; PMID:8431945
A:Accession: A45174
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1760 <BON>
A:Cross-references: GB:U08501; NID:g157975; PID:g157976
A:Experimental source: adult head
A:Note: sequence extracted from NCBI backbone (NCBI:P124850)
A:Genetics:
A:Gene: FlyBase:cli; eya; eyes absent
A:Cross-references: FlyBase:FBgn0000320

Query Match 8.8%; Score 96; DB 2; Length 760;
Best Local Similarity 23.7%; Pred. No. 2.9;
Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6;

QY 42 GSSMY-----SGSSSLVYTSAMPGFYPATATAAAYGALRGKRVYNTFRAAA 92
DB 223 GSNLYGSSASNPIDGGAIVAVNSSAV-----AAAAAAVYDGK-----DYYYSMOOYT 272
QY 93 PPPPIPAYGVVYQEPYVGNKLLQGYAAVRYAQTPTATAA----- 134
DB 273 PPPPYSGYGTPTAAATARQAKMEGAAAAAAYLTPTSYAASGNNNSQLYSSFYAGYNNF 332
QY 135 -----YSDSYGRVYAADPYHHTLAPAP-----TYGVG-AMNAFAPLTDKTRSHA 178
DB 333 GQODYGVYNGQYNNYSPANYSPYAVSSPSSASHGHGFVAASNSNSESPDTJHS 389

RESULT 12

T15142
hypothetical protein T28F2.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C:Accession: T15142
R:Madson, C.; Fironick, B.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid T28F2.
A:Reference number: Z18300
A:Accession: T15142
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-418 <MAD>
A:Cross-references: EMBL:AF00198; NID:G2047345; PID:G2047346; PIDN:AAB53052.1; GSPDB:GN
C:Experimental source: strain Bristol N2; clone T28F2
C:Genetics:
A:Gene: CESP:T28F2.6
A:Map position: 1
A:Intons: 49/3
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 8.7%; Score 95; DB 2; Length 418;
Best Local Similarity 30.7%; Pred. No. 1.8;
Matches 42; Conservative 7; Mismatches 60; Indels 28; Gaps 5;

QY 40 QEGSSMYSGPSLVYTSAMP-----FPYPATATAAAYGALRGGR 82
DB 272 KDGAGPDGPAGAGCTDAGPDDAAYCCPPRTTGAGAYPRGGAAAAAPAGYDGGAGA 331
QY 83 TVYNTFRAAAPPPPIPAYGVVYQEPYVGNKLLQGYAAVRYAQTPTATAAAYSDSYGRV 142
DB 332 APEAAPAAAAAPAPAPAAAA--APAG--YQCG-AAAGAAAPPPAPAAAAAEPAPAPA 385
QY 143 YAADPYHHTLAPAPYVG 159
DB 386 PAAAP-----PPAPAG 397

RESULT 13
T20047
hypothetical protein C49A1.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20047
R:Mathews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z19217
A:Accession: T20047
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-469 <MTL>
A:Cross-references: EMBL:Z83221; PIDN:CAB05707.1; GSPDB:GN00019; CESP:C49A1.4
A:Experimental source: clone C49A1
C:Genetics:
A:Gene: CESP:C49A1.4
A:Map position: 1
A:Intons: 37/1; 98/1; 140/3; 249/3; 318/1; 441/3

Query Match 8.7%; Score 94.5; DB 2; Length 469;
Best Local Similarity 23.6%; Pred. No. 2.2;
Matches 51; Conservative 22; Mismatches 80; Indels 63; Gaps 9;

QY 18 PVGAVYSPDFYAGTVLLCOANQEGSSMY-----SGPSSLVYTSAMP----- 59
DB 25 PTVDLAISEAYGSSSTSTSLTSSVTQYNSYPOYAMTSAHPANYVQGVNTANLACT 84
QY 60 -GPPYPATATAAAYGALRGKRVYNTFRAAAPPPPIPAYGVVYQEPYVGNKLLQGG 118
DB 85 TAFPYSLTT--PSYGSY-----PVDYTSAAAA-----YONPYTN--LRGG 122
QY 119 YAAVRYAQTPTATAAAYSDSYGRVYAADPYHHTLAPAPYGV--GAMNAFA-----PLTDA 172
DB 123 TAAVYVNPINATTAATAAAYSAVSVLGTDAVNLGTSDSGSGVPSVTISFSLKEKPKVSK 182
QY 173 KTRSHA-----DDVGLVLSLQASI 192
DB 183 KKTGSCSGPDERTARVFMIDIDIAVISRNLYASV 218

RESULT 14

S42674
adhesive protein - bifurcate mussel (fragments)
N:Alternate names: polyphenolic protein
C:Species: *Septifer bifurcatus* (bifurcate mussel)
C>Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 07-May-1999
C:Accession: S42674
R:Repecki, L.M.; Chiu, S.S.; Waite, J.H.; Lavin, M.F.
Mol. Marine Biol. Biotechnol. 1, 78-88, 1991
A>Title: Molecular diversity of marine glues: polyphenolic proteins from five mussel spec
A:Reference number: S42671; MUID:93364572; PMID:1845474
A:Accession: S42674
A:Molecule type: protein
A:Residues: 1-13;14-36;37-39;40-52;53-65;66-78;79-81;82-86;87-94;95-101;102-109;110-117;1
A:Note: Ser-19 was also found

Query Match 8.7%; Score 94; DB 2; Length 219;
Best Local Similarity 28.7%; Pred. No. 1;
Matches 39; Conservative 14; Mismatches 47; Indels 36; Gaps 9;

QY 46 YSGPSSLVYTSAMPGFYPATATAAAYGALRG-----RTVYNT--FRAAP 93
DB 64 YKYPTR--YTYGTGYVPAKAPAKTSYGTGYKSYGVYKSPYGTGYKSSYGTGYKAPAK 121
QY 94 PPPPIPAYGVVYQEPV-----YGNKLLQGYAAVRYAQTPTATAAAYSDSYGRVYAA 145
DB 122 PPSKPSYGTGYKAPKTYTKTPSSYGT-----GYKA-----PTKYTKK--PSSYGTGYKA 169
QY 146 DPYHHTLAPAPYVG 161
DB 170 -PTKYSTKPS--SYGTG 183

RESULT 15
H86334
T20H2.10 protein - *Arabidopsis thaliana*

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 21:21:40 ; Search time 16 Seconds
(without alignments)

526.231 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTNRKAVNPYTNMGKLNPPV.....VLSSLQASIVQGGYNRFAPY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020.5	94.0	396	1 A2BP_MOUSE	Q9J343 mus musculu
2	1009.5	93.0	397	1 A2BP_HUMAN	Q9HWD1 homo sapien
3	307.5	28.3	391	1 RBM9_HUMAN	O43251 homo sapien
4	116	10.7	363	1 CALIC_HUMAN	O99715 homo sapien
5	115	10.6	3119	1 CALIC_MOUSE	O60847 mus musculu
6	103	9.5	3124	1 CALIC_CHICK	P13944 gallus gall
7	98.5	9.1	376	1 FXE1_HUMAN	O00338 homo sapien
8	98.5	9.1	655	1 EMS_MOUSE	O61545 mus musculu
9	98	9.0	376	1 HMEV_DROME	O63252 rattus norv
10	97	8.9	497	1 BRN1_RAT	O05201 drosophila
11	96	8.8	766	1 EYA_DROME	O62424 mus musculu
12	94	8.7	386	1 HKAD_MOUSE	P31271 homo sapien
13	94	8.7	388	1 HKAD_HUMAN	P17110 ceratilis c
14	93.5	8.6	320	1 CH36_CERCA	P33693 gallus gall
15	93	8.6	387	1 GAT6_CHICK	P31361 mus musculu
16	93	8.6	495	1 BRN1_MOUSE	O01844 homo sapien
17	92.5	8.5	656	1 EMS_HUMAN	P20264 homo sapien
18	92	8.5	500	1 BRN1_HUMAN	Q92754 homo sapien
19	90.5	8.3	450	1 AP2C_HUMAN	P56645 homo sapien
20	90.5	8.3	1210	1 Y102_MYCTU	O53951 mycobacteri
21	89	8.2	463	1 NK2C_MOUSE	P97334 mus musculu
22	87.5	8.1	362	1 NK2C_HUMAN	O16676 homo sapien
23	87	8.0	465	1 FXD1_HUMAN	P08699 rattus norv
24	86.5	8.0	261	1 LEG3_RAT	Q9Z2V7 drosophila
25	86	7.9	568	1 ATPA_DEIRA	P54320 mus musculu
26	86	7.9	860	1 ELS_MOUSE	Q27002 toxoplasma
27	85.5	7.9	345	1 GR44_TOXGO	P18111 mus musculu
28	85	7.8	268	1 CDX1_MOUSE	P00762 bacillus am
29	85	7.8	382	1 SUBT_BACAM	P16376 drosophila
30	85	7.8	746	1 7UP2_DROME	P20469 pantoea ana
31	85	7.8	1322	1 ICEA_PAVAN	P81445 alcaigenes
32	84.5	7.8	330	1 NIR_ALCXX	P07518 bacillus pu
33	84	7.7	275	1 SUBT_BACPU	

34	84	7.7	876	1 BGL1_SACFI	P22506 saccharomyc
35	83.5	7.7	466	1 ANX7_HUMAN	P20073 homo sapien
36	83.5	7.7	520	1 LAC1_TRAVI	O99044 trameetes vi
37	83.5	7.7	741	1 TLE4_RAT	O07141 rattus norv
38	83.5	7.7	766	1 TLE4_HUMAN	O04727 homo sapien
39	83.5	7.7	766	1 TLE4_MOUSE	O62441 mus musculu
40	83	7.6	450	1 FAX8_HUMAN	O06710 homo sapien
41	83	7.6	463	1 ANX7_MOUSE	O07076 mus musculu
42	83	7.6	621	1 VP40_HSVBC	P54817 bovine herp
43	82.5	7.6	520	1 LAC1_CORHI	O02497 coriolus hi
44	82.5	7.6	538	1 SNPH_HUMAN	O15079 homo sapien
45	82.5	7.6	604	1 GLAS_DROME	P13360 drosophila

ALIGNMENTS

RESULT 1
A2BP_MOUSE STANDARD; PRT; 396 AA.
ID A2BP_MOUSE
AC Q9J343; Q9JTB5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ataxin 2-binding protein.
GN A2BP1 OR A2BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Osada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 21-396 FROM N.A.
RX MEDLINE=2075619; PubMed=10814712;
RA Shibata H., Huynh D.P., Pulst S.-M.;
RL "A novel protein with RNA-binding motifs interacts with ataxin-2."
Hum. Mol. Genet. 9:1303-1313(2000).
CC -1- SUBMIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB041596; BAA95079.1; -;
DR EMBL: AF107204; AAF78292.1; -;
DR HSPB; P11940.1; CVT
DR MGD; MGI:1926224; A2bp1.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PSS0030; RRM_RNP_1; 1.
KW RNA-binding.
FT DOMAIN 116 192 RNA-BINDING (RRM).
FT CONFLICT 60 G -> D (IN REF. 2).
FT CONFLICT 128 R -> RFR (IN REF. 2).
FT CONFLICT 369 G -> S (IN REF. 2).
SQ SEQUENCE 396 AA; 42590 MM; 9A0C59C5E687F39F CRC64;
Query Match 94.0%; Score 1020.5; DB 1; Length 396;
Best Local Similarity 94.6%; Pred. No. 1.8e-74;

Matches 192; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 MTNKKAVNPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOESSSMYSGGSSLVYTSAMP 60
 DB 195 MTNKKAVNPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOESSSMYSGGSSLVYTSAMP 254

QY 61 PPYPATAAAARGAHLRGGRGRTVNTFPRAAAPPPPIPAVGGVYOEFPYGNKLLQGGYA 120
 DB 255 PPYPATAAAARGAHLRGGRGRTVNTFPRAAAPPPPIPAVGGVYODGFYGD-1YGGYA 313

QY 121 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAFAPLTDKTSHAD 180
 DB 314 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAFAPLTDKTSHAD 373

QY 181 VGLVLSLQASLYOGGYNRFAPY 203
 DB 374 VGLVLSLQASLYOGGYNRFAPY 396

RESULT 2

A2BP_HUMAN STANDARD; PRT; 397 AA.

AC Q9NWB1. Q9NS20;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ataxin-2-binding protein.

OS Homo sapiens (Human).

GN A2BP1 OR A2BP.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RA Nucleic acid sequence from N.A.

RA Isegaki T., Oca T., Hayashi K., Sugiyama T., Otsubo T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,

RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takehachi M., Chiba Y., Ishida S., Murakawa K., Ono W., Takiguchi S.,

RA Takehachi S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,

RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,

RA Nimomura K., Iwayanagi T.,

RT "NEDD human cDNA sequencing project."

RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE OF 21-397 FROM N.A.

RX MEDLINE=20275619; PubMed=10814712;

RA Shibata H., Huynh D.P., Pulst S.-M.,

RL "A novel protein with RNA-binding motifs interacts with ataxin-2";

CC Hum. Mol. Genet. 9:1303-1313(2000).

CC -1-SUBUNIT BINDS TO THE C-TERMINAL OF ATAXIN-2.

CC -1-TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MUSCLE AND BRAIN.

CC -1-SIMILARITY: CONTRAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AK001027; BAA91472.1; -

DR EMBL; AF107203; AAF78291.1; -

DR HSSP; P11940; ICVJ.

DR MIM; 605104; -

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF00076; rrm; 1.

DR SMART; SM00360; RRM; 1.

DR PROSITE; PS50102; RRM; 1.

DR PROSITE; PS00030; RRM_RNP_1; 1.

KM RNA-binding.

FT DOMAIN 117 193 RNA-BINDING (RRM).

FT CONFLICT 92 92 A->T (IN REF. 2).

SQ SEQUENCE 397 AA; 42754 MW; E3E9060B68C79880 CRC64;

Query Match 93.0%; Score 1009.5; DB 1; Length 397;

Best Local Similarity 93.6%; Pred. No. 1.4e-73;

Matches 190; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 1 MTNKKAVNPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOESSSMYSGGSSLVYTSAMP 60

DB 196 MTNKKAVNPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOESSSMYSGGSSLVYTSAMP 255

QY 61 PPYPATAAAARGAHLRGGRGRTVNTFPRAAAPPPPIPAVGGVYOEFPYGNKLLQGGYA 120

DB 256 PPYPATAAAARGAHLRGGRGRTVNTFPRAAAPPPPIPAVGGVYODGFYGD-1YGGYA 314

QY 121 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAFAPLTDKTSHAD 180

DB 315 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAFAPLTDKTSHAD 374

QY 181 VGLVLSLQASLYOGGYNRFAPY 203

DB 375 VGLVLSLQASLYOGGYNRFAPY 397

RESULT 3

RM9_HUMAN STANDARD; PRT; 391 AA.

AC Q43251. Q9UGW4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative RNA-binding protein 9 (RNA binding motif protein 9).

GN RBM9.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RA Nucleic acid sequence from N.A.

RA Whiteley M.,

RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.

CC -1-SIMILARITY: CONTRAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AL009266; CAAL5842.1; ALT_SEQ.

DR EMBL; AL049748; CAB63054.1; -

DR HSSP; P11940; ICVJ.

DR GeneW; HGNC:9906; RBM9.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF00076; rrm; 1.

DR SMART; SM00360; RRM; 1.

DR PROSITE; PS50102; RRM; 1.

DR PROSITE; PS00030; RRM_RNP_1; 1.

KM RNA-binding.

FT DOMAIN 112 188 RNA-BINDING (RRM).

FT SEQUENCE 391 AA; 42284 MW; ED2B76C03D16A7B CRC64;

Query Match 28.3%; Score 307.5; DB 1; Length 391;

Best Local Similarity 43.0%; Pred. No. 1.3e-17;

Matches 77; Conservative 13; Mismatches 56; Indels 31; Gaps 7;

QY 1 MTNKKAVNPYTNMGKLNIPVGAIVSPDFYAGTVLLCOAN--OEGSSMYSGGSSLVYTSAMP 56

```

Db      191 MTNKKWVTPYANGWKLSPVGVAVGVPELHYAASSFOADVSLGNDAAVPLSGGQJINTYIPL 250
Qy      57 AMGEFVP-AATAAAYAGAHGRGRVTVTFRAAAPPPPIPGVYVDEPVYGNLL 115
Db      251 IIPGPPTPAATTAAGAHGRGRVTVGAVR-ADPPTAIIPALPGVYVDDGFYADLY 309
Qy      116 QGGAAYRY---AQPTPATAAAYSXGYVADPHTHTLAPATYGVGMNAPAPLT 170
Db      310 IESANCRSNRNVDMQPT-----DMHSLLLQPPQ---LLQPLQPLT 347

RESULT 4
CAIC_HUMAN STANDARD; PRT; 3063 AA.
AC      099715; 099716;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DE      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Collagen alpha 1(XII) chain precursor.
GN      COL12A1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A. (LONG AND SHORT FORM), AND PARTIAL SEQUENCE.
RX      MEDLINE=97288521; PubMed=9143499;
RA      Gerecke D.R., Olson P.F., Koch W., Knoll J.H.M., Taylor R.,
RA      Hudson D.L., Champilaud M.-F., Olsen B.R., Burgess R.E.;
RT      "Complete primary structure of two splice variants of collagen XII,
RT      and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(XI)
RT      collagen (COL19A1), and alpha 1(XIX) collagen (COL19A1) to human
RT      chromosome 6q12-q13."
RL      Genomics 41:236-242(1997).
CC      -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC      CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC      SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC      LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC      -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDA OF
CC      NONTRIPLE-HELICAL SEQUENCES.
CC      -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC      SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC      FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC      OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC      ISOFORM CHAINS.
CC      -1- TISSUE SPECIFICITY: FOUND IN COLLAGEN I-CONTAINING TISSUES: BOTH
CC      SHORT AND LONG ISOFORMS APPEAR IN AMNION, CHORION, SKELETAL
CC      MUSCLE, SMALL INTESTINE, AND IN CELL CULTURE OF DERMAL
CC      FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT
CC      ISOFORM IS FOUND IN LONG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL
CC      CARCINOMA CELL LINE.
CC      -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC      EACH END (BY SIMILARITY).
CC      -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC      SIMILARITY).
CC      -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY THE
CC      LONG VARIANT IS A PROTEOGLYCAN.
CC      -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC      INTERRUPTED HELICES (FACIT) FAMILY.
CC      -1- SIMILARITY: CONTAINS 4 VWFA DOMAINS.
CC      -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U73778; AAC51244.1; -

```

```

DR      EMBL; U73779; AAD40483.1; -.
DR      HSSP; P02751; 1TF.
DR      Genew; HGNC:2188; COL12A1.
DR      MIM; 120320; -.
DR      InterPro; IPR000087; Collagen.
DR      InterPro; IPR003961; FN_III.
DR      InterPro; IPR003962; FNIII_repeat.
DR      InterPro; IPR003129; TSPN.
DR      InterPro; IPR02035; VWF_A.
DR      Pfam; PF00041; Fn3; 18.
DR      Pfam; PF00092; Wnt4; 4.
DR      Pfam; PF01391; Collagen; 4.
DR      Pfam; PF02210; TSPN; 1.
DR      PRINTS; PR00014; FNTYDEIIL.
DR      PRINTS; PR00453; VWFADOMAIN.
DR      SMART; SM00060; FN3; 17.
DR      SMART; SM00210; TSPN; 1.
DR      SMART; SM00327; VWF; 4.
DR      PROSITE; PS50234; VWFA; 4.
KW      Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW      Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
FT      SIGNAL 1 24
FT      CHAIN 25 3063 COLLAGEN ALPHA 1(XII) CHAIN.
FT      DOMAIN 25 114 FIBRONECTIN TYPE-III 1.
FT      DOMAIN 140 316 VWFA 1.
FT      DOMAIN 333 426 FIBRONECTIN TYPE-III 2.
FT      DOMAIN 440 616 VWFA 2.
FT      DOMAIN 630 721 FIBRONECTIN TYPE-III 3.
FT      DOMAIN 722 812 FIBRONECTIN TYPE-III 4.
FT      DOMAIN 813 903 FIBRONECTIN TYPE-III 5.
FT      DOMAIN 904 998 FIBRONECTIN TYPE-III 6.
FT      DOMAIN 999 1085 FIBRONECTIN TYPE-III 7.
FT      DOMAIN 1086 1178 FIBRONECTIN TYPE-III 8.
FT      DOMAIN 1199 1371 VWFA 3.
FT      DOMAIN 1384 1473 FIBRONECTIN TYPE-III 9.
FT      DOMAIN 1474 1564 FIBRONECTIN TYPE-III 10.
FT      DOMAIN 1565 1652 FIBRONECTIN TYPE-III 11.
FT      DOMAIN 1654 1751 FIBRONECTIN TYPE-III 12.
FT      DOMAIN 1752 1842 FIBRONECTIN TYPE-III 13.
FT      DOMAIN 1843 1932 FIBRONECTIN TYPE-III 14.
FT      DOMAIN 1933 2023 FIBRONECTIN TYPE-III 15.
FT      DOMAIN 2024 2114 FIBRONECTIN TYPE-III 16.
FT      DOMAIN 2115 2202 FIBRONECTIN TYPE-III 17.
FT      DOMAIN 2203 2291 FIBRONECTIN TYPE-III 18.
FT      DOMAIN 2323 2496 VWFA 4.
FT      DOMAIN 2451 2746 NONHELICAL REGION (NC3)
FT      DOMAIN 2747 2898 TRIPLE-HELICAL REGION (COL2)
FT      DOMAIN 2899 2941 WITH 1 IMPERFECTION.
FT      DOMAIN 2942 3044 NONHELICAL REGION (NC2).
FT      DOMAIN 3045 3063 TRIPLE-HELICAL REGION (COL1)
FT      BINDING 798 889 WITH 2 IMPERFECTIONS.
FT      BINDING 889 981 NONHELICAL REGION (NC1).
FT      SITE 862 984 TO CHONDROITIN SULFATE (POTENTIAL).
FT      SITE 2779 2781 CELL ATTACHMENT SITE (POTENTIAL).
FT      SITE 2895 2897 CELL ATTACHMENT SITE (POTENTIAL).
FT      MOD_RES 2895 2944 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2944 2947 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2947 2950 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2950 2959 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2959 2965 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2965 2966 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2966 2968 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2971 2971 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2971 2983 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 2983 3000 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 3000 3003 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 3003 3014 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 3014 3023 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 3023 3026 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 3026 3029 HYDROXYLATION (BY SIMILARITY).
FT      MOD_RES 3029 3029 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 700

```

```

FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2206 2206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2528 2528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 25 1188 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 3063 AA; 333189 MW; 75FEA78FA8E48293 CRC64;

Query Match 10.7%; Score 116; DB 1; Length 3063;
Best Local Similarity 25.1%; Pred. No. 0.21;
Matches 45; Conservative 16; Mismatches 54; Indels 64; Gaps 8;

Oy 50 SLVVTSMAMPGRPPYPAATAAAYR---GAHLGRGRTV-----YNTFRAA 92
Db 2078 NVVLIQPLQPDPPY-KITVIAVEDGDGHLTGNGRTVGLPPONIHISDEWTRFRVSW 2136
Oy 93 PPPPIPAVGAVYQEPVNVKLLQ---GGVAAVRYAQTPATAAVSDSGRYAADPYH 149
Db 2137 DDPSPVLYCKIVYKRVGSENEPMEAVGEMTSTYL----- 2171
Oy 150 HTLAPAPTYGVGAMNAFA-----PLTDKTRSHADVGLVLSLQASIVQGYNRP 200
Db 2172 HMLNSTTVDV---NVYAQYDSGLSVPLTDQGT-----LYLVNTDLKTYQIGMDTF 2220

RESULT 5
CAIC MOUSE STANDARD; PRT; 3119 AA.
ID CAIC MOUSE STANDARD; PRT; 3119 AA.
AC 060847; P70322;
AC 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN COL12A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-1 AND XI1B-1).
RC STRAIN=Swiss Webster, and C57BL/6J; TISSUE=Skin;
RX MEDLINE=96170761; PubMed=86101036;
RA Boelme K., Li Y., Oh P.S., Olsen B.R.;
RA Kania B.R., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W.,
RA Olsen B.R., Nishimura I.;
RT "Structural variation of type XII collagen at its carboxyl-terminal NCI domain generated by tissue-specific alternative splicing.";
RT J. Biol. Chem. 274:22053-22059(1999).
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF NONTRIPE-HELICAL SEQUENCES (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: XI1A-1 (SHOWN HERE), XI1A-2/ERNA, XI1B-1 AND XI1B-2, ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OR ANY COMBINATION OF THE VARIOUS ISOFORMS.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHOONDRIUM, SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERIOSTEUM.
CC -1- DEVELOPMENTAL STAGE: THE LONG NC3 XI1A ISOFORMS ARE PREDOMINANT AT EARLY STAGES (ED7 AND 11); AT LATER STAGES OF DEVELOPMENT (ED15 AND 17) THE SHORT NC3 XI1B FORMS BECOME THE MAJOR FORMS. AS THE SHORT NC3 FORMS BECOME THE MAJOR PRODUCT, THE LONG SPLICED VARIANT

```

```

CC CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER BIRTH.
CC THE LONG NC1 ISOFORMS, XI1A-1 AND XI1B-1, PEAK IN 15-DAY OLD EMBRYOS AND DECREASE IN 17-DAY OLD ONES. THE EXPRESSION OF THE SHORT NC1 FORM XI1B-2 REMAINS CONSTANT THROUGHOUT LATE STAGES OF EMBRYONIC DEVELOPMENT (ED15 AND ED17).
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END (BY SIMILARITY).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY SIMILARITY).
CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY ISOFORM XI1A-2 IS A PROTEOGLYCAN.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 4 WMFA DOMAINS.
CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U25652; AAA9719.1; ALT_SEQ.
CC EMBL; U57095; AAB07047.1; -.
CC HSSP; P02751; IFNA.
CC MGD; MGI:88448; Col12a1.
CC InterPro; IPR000087; Col12a1.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003962; FNIII_repeat.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR002035; WMFA.
CC Pfam; PF00041; Fn3; 18.
CC Pfam; PF00092; Wfa; 4.
CC Pfam; PF01391; Collagen; 4.
CC Pfam; PF02210; TSPN; 1.
CC PRINTS; PR00014; FNTYPEIII.
CC PRINTS; PR00453; WMFADOMAIN.
CC SMART; SMO0060; FN3; 16.
CC SMART; SMO0210; TSPN; 1.
CC SMART; SMO0327; WFA; 4.
CC PROSITE; PS50234; WMFA; 4.
CC KX Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
CC KW SIGNAL.
CC FT 1 24
CC FT CHAIN 25 3119
CC FT DOMAIN 25 114 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 140 316 WMFA 1.
CC FT DOMAIN 332 423 FIBRONECTIN TYPE-III 2.
CC FT DOMAIN 444 620 WMFA 2.
CC FT DOMAIN 634 725 FIBRONECTIN TYPE-III 3.
CC FT DOMAIN 726 816 FIBRONECTIN TYPE-III 4.
CC FT DOMAIN 817 907 FIBRONECTIN TYPE-III 5.
CC FT DOMAIN 908 1002 FIBRONECTIN TYPE-III 6.
CC FT DOMAIN 1003 1089 FIBRONECTIN TYPE-III 7.
CC FT DOMAIN 1090 1182 FIBRONECTIN TYPE-III 8.
CC FT DOMAIN 1203 1375 WMFA 3.
CC FT DOMAIN 1388 1474 FIBRONECTIN TYPE-III 9.
CC FT DOMAIN 1480 1568 FIBRONECTIN TYPE-III 10.
CC FT DOMAIN 1569 1652 FIBRONECTIN TYPE-III 11.
CC FT DOMAIN 1659 1757 FIBRONECTIN TYPE-III 12.
CC FT DOMAIN 1758 1848 FIBRONECTIN TYPE-III 13.
CC FT DOMAIN 1849 1938 FIBRONECTIN TYPE-III 14.
CC FT DOMAIN 1939 2029 FIBRONECTIN TYPE-III 15.
CC FT DOMAIN 2030 2120 FIBRONECTIN TYPE-III 16.
CC FT DOMAIN 2121 2208 FIBRONECTIN TYPE-III 17.
CC FT DOMAIN 2209 2297 FIBRONECTIN TYPE-III 18.
CC FT DOMAIN 2329 2501 WMFA 4.
CC FT DOMAIN 2456 2751 NONHELICAL REGION (NC3).
CC FT DOMAIN 2752 2899 TRIPLE-HELICAL REGION (COL2) WITH 1 IMPERFECTION.

```



```

FT DOMAIN 2900 2942 NONHELICAL REGION (NC2).
FT DOMAIN 2943 3045 TRIPLE-HELICAL REGION (COL1)
FT 3046 3119 WITH 2 IMPERFECTIONS.
FT DOMAIN 802 802 NONHELICAL REGION (NC1).
FT BINDING 893 893 TO CHONDROITIN SULFATE (POTENTIAL).
FT BINDING 985 985 TO CHONDROITIN SULFATE (POTENTIAL).
FT SITE 866 868 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2784 2786 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2896 2898 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 2945 2945 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2948 2948 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2951 2951 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2960 2960 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2966 2966 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2969 2969 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2972 2972 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2984 2984 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3001 3001 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3004 3004 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3015 3015 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3024 3024 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3027 3027 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3030 3030 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 869 872 POLY-THR.
FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2212 2212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2684 2684 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 25 1190 MISSING (IN ISOFORM XIIB-1 AND ISOFORM
FT XIIB-2).
FT VARSPIC 3062 3064 EPI -> GSG (IN ISOFORM XIIB-2 AND ISOFORM
FT XIIB-2).
FT VARSPIC 3065 3119 MISSING (IN ISOFORM XIIB-2 AND ISOFORM
FT XIIB-2).
SQ SEQUENCE 3119 AA; 340239 NM; 981F99C86A3251 CRC64;
Query Match 10.6%; Score 115; DB 1; Length 3119;
Best Local Similarity 24.4%; Pred. No. 0.26;
Matches 43; Conservative 16; Mismatches 59; Indels 58; Gaps 7;
QY 50 SLYVTSAMPSPYPAATAAAYR---GAHLGRGRV-----YTFPRAA 92
DB 2084 NNVLILQLPQPTPY-KIVIAIYEDGDGHLGNRTYGLAPQNIHIFEDENYCFRFSW 2142
QY 93 PPPPIPAYGVVYQEPVYGNKLLQ---GQYAAVRYAQPTPATAAAYSISYGRVYAADPYH 149
DB 2143 DPSPSPVLYGYKIVYKPGVSNPEMFAVGEVTSYTL----- 2177
QY 150 HTLAPAPYVGAMNAF-----APLTDKTRSHADVDGLVLSLQASYOGGVNRF 200
DB 2178 HNLNPSTTVDSVYAQYDGLSVPLTDGTT-----LYLNVTDLKYQVGWMTF 2226

```

```

RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA Nishida Y., Obata M., Kinata K.;
RT "The complete primary structure of type XII collagen shows a chimeric
RT molecule with reiterated fibronectin type III motifs, von Willebrand
RT factor A motifs, a domain homologous to a noncollagenous region of
RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT site." J. Cell Biol. 115:209-221(1991).
RN [2]
RN SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP 2846-2873.
RX MEDLINE=90062079; Pubmed=2584192;
RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
RT "Type XII collagen. A large multidomain molecule with partial
RT homology to type IX collagen." J.
RL J. Biol. Chem. 264:19772-19778(1989).
RN [3]
RN SEQUENCE OF 2960-3076 FROM N.A.
RX MEDLINE=97317590; Pubmed=3476925;
RA Gordon M.K., Gerecke D.R., Olsen B.R.;
RT "Type XII collagen: distinct extracellular matrix component
RT discovered by cDNA cloning." J.
RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
RN [4]
RN SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.
RP TISSUE=Embryo;
RX MEDLINE=93042014; Pubmed=1420368;
RA Trueb J., Trueb B.;
RT "The two splice variants of collagen XII share a common 5' end." J.
RL Biochim. Biophys. Acta 1171:97-98(1992).
RN [5]
RN ALTERNATIVE SPLICING.
RX MEDLINE=95370352; Pubmed=7642694;
RA Koch M., Bohmann B., Matchison M., Hagios C., Trueb B., Chiquet M.;
RT "Large and small splice variants of collagen XII: differential
RT expression and ligand binding." J.
RL J. Cell Biol. 130:1005-1014(1995).
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX.
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDA OF
CC NONTRIPLE-HELICAL SEQUENCES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYTICALLY THE LARGE
CC ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN
CC THE SMALL.
CC -1- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,
CC LIGAMENTS, PERICHONDRUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE
CC TISSUES CONTAINING TYPE I COLLAGEN.
CC -1- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS. TWO TRIPLE-
CC HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL
CC DOMAINS (NC1, NC2, AND NC3).
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 4 WMPA DOMAINS.
CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; D00824; BAA0701.1; -
DR EMBL; X61024; CAA4358.1; -
DR EMBL; M17375; AAA48718.1; -
DR EMBL; J05137; AAA48635.1; -
DR EMBL; X67327; CAA47744.1; -
DR PIR; A28037; A28037.
DR PIR; A34485; A34485.
DR HSSP; P20701; ILPA.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00092; vwa; 4.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF02210; TSPN; 1.
DR PRINTS; PR00014; FNYPEI11.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS50234; VWA; 4.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 3124
FT DOMAIN 25 114
FT DOMAIN 139 311
FT DOMAIN 332 425
FT DOMAIN 439 615
FT DOMAIN 629 720
FT DOMAIN 721 811
FT DOMAIN 812 904
FT DOMAIN 905 998
FT DOMAIN 999 1085
FT DOMAIN 1086 1178
FT DOMAIN 1179 1371
FT DOMAIN 1384 1473
FT DOMAIN 1474 1565
FT DOMAIN 1566 1654
FT DOMAIN 1655 1755
FT DOMAIN 1756 1846
FT DOMAIN 1847 1936
FT DOMAIN 1937 2027
FT DOMAIN 2028 2118
FT DOMAIN 2119 2205
FT DOMAIN 2207 2295
FT DOMAIN 2297 2387
FT DOMAIN 2387 2455
FT DOMAIN 2455 2750
FT DOMAIN 2751 2902
FT DOMAIN 2903 2945
FT DOMAIN 2946 3048
FT DOMAIN 3049 3124
FT DOMAIN 3086 3096
FT DOMAIN 3113 3123
FT BINDING 797 797
FT BINDING 890 890
FT BINDING 981 981
FT SITE 2448 2440
FT SITE 2899 2901
FT CARBOHYD 32 32
FT CARBOHYD 1006 1006
FT CARBOHYD 1032 1032
FT CARBOHYD 1044 1044
FT CARBOHYD 1512 1512
FT CARBOHYD 1767 1767
FT CARBOHYD 2210 2210
FT CARBOHYD 2273 2273
FT CARBOHYD 2532 2532
-----
COLLAGEN ALPHA 1(XII) CHAIN.
FIBRONECTIN TYPE-III 1.
VWFA 1.
FIBRONECTIN TYPE-III 2.
VWFA 2.
FIBRONECTIN TYPE-III 3.
FIBRONECTIN TYPE-III 4.
FIBRONECTIN TYPE-III 5.
FIBRONECTIN TYPE-III 6.
FIBRONECTIN TYPE-III 7.
FIBRONECTIN TYPE-III 8.
VWFA 3.
FIBRONECTIN TYPE-III 9.
FIBRONECTIN TYPE-III 10.
FIBRONECTIN TYPE-III 11.
FIBRONECTIN TYPE-III 12.
FIBRONECTIN TYPE-III 13.
FIBRONECTIN TYPE-III 14.
FIBRONECTIN TYPE-III 15.
FIBRONECTIN TYPE-III 16.
FIBRONECTIN TYPE-III 17.
FIBRONECTIN TYPE-III 18.
VWFA 4.
NONHEMICAL REGION (NC3).
TRIPLE-HELICAL REGION (COL2)
WITH 1 IMPERFECTION.
NONHEMICAL REGION (NC2).
TRIPLE-HELICAL REGION (COL1)
WITH 2 IMPERFECTIONS.
NONHEMICAL REGION (NC1).
ASP/GLU-RICH (ACIDIC).
ARG/LYS-RICH (BASIC).
TO CHONDROITIN SULFATE (POTENTIAL).
TO CHONDROITIN SULFATE (POTENTIAL).
CELL ATTACHMENT SITE (POTENTIAL).
CELL ATTACHMENT SITE (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 2683 2683 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VASPLIC 25 1188 MISSING (IN SHORT ISOFORM).
FT CONFLICT 1258 1258 T -> S (IN REF. 4).
FT CONFLICT 1264 1264 D -> E (IN REF. 4).
FT CONFLICT 2759 2759 P -> A (IN REF. 2).
FT CONFLICT 2803 2803 L -> F (IN REF. 2).
FT CONFLICT 2977 2977 V -> F (IN REF. 2).
FT CONFLICT 3075 3076 OP -> AG (IN REF. 3).
SQ SEQUENCE 3124 AA; 340578 MW; 094285AFEF346CF CRC64;

Query Match 9.5%; Score 103; DB 1; Length 3124;
Best Local Similarity 25.1%; Pred. No. 2.3;
Matches 42; Conservative 15; Mismatches 46; Indels 64; Gaps 8;

QY 62 PYPAATAAAYR---GAHLRGGRIV-----YTFPAAPPPPIPAYGVV 104
DB 2094 PY-KITVAAYVEDGGGGLTGNGRTGLPPQNIYITDEWYFRFVSDPSPVILGYKI 2152
QY 105 YOEPIYGNKLDL---GGAAYRYAQPRTATAAAYSDSGRYVAADPHHTLAPPTGVG 161
DB 2153 VYKPVGSNEPMVEFVGEVTSYTL-----HNLSPTTYDV- 2186
QY 162 ANNAFA-----PLTDAKTRSHADVGLVLSLQASIIYGGYNRF 200
DB 2187 --NVTAQDMSGMSITLDGTT-----LYNVATDITTYKIGMDTF 2224

RESULT 7
FXEL HUMAN STANDARD; PRT; 376 AA.
AC 000158; 075765;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Forkhead box protein E1 (Forkhead-related protein FKHL15) (Thyroid
DE transcription factor 2) (TFP-2).
GN FOXE1 OR FKHL15 OR TTF2 OR TTF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=retinocytes;
RX MEDLINE=97312696; PubMed=9169137;
RA Chadwick B.P.; Obermayr F.; Frischauf A.-M.;
RT "FKHL15, a new human member of the forkhead gene family located on
RT chromosome 9q22.4";
RL Genomics 41:390-396(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Macchia P.E.; Mattei M.-G.; Lapi P.; Penzi G.; Di Lauro R.;
RT "Cloning, chromosomal localization and identification of two
RT polymorphisms in the human thyroid transcription factor 2 gene
RT (TFP2).";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP VARIANT BAMFORTH-LAZARUS SYNDROME VAL-65.
RX MEDLINE=98361170; PubMed=9697705;
RA Clifton-Bligh R.J.; Wentworth J.M.; Heinz P.; Crisp M.S.; John R.;
RA Lazarus J.H.; Ludgate M.; Charterjee V.K.;
RT "agenesis, cleft palate and choanal atresia.";
RL Nat. Genet. 19:399-401(1998).
CC -I- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. COULD BE INVOLVED IN
CC THYROID GLAND ORGANOGENESIS.
CC -I- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -I- TISSUE SPECIFICITY: DETECTED IN ADULT BRAIN, PLACENTA, LUNG,
CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS, HEART, COLON, SMALL
CC INTESTINE TESTIS AND THYMUS. EXPRESSION WAS STRONGEST IN HEART AND
CC PANCREAS.
CC -I- DISEASE: DEFECTS IN FOXE1 ARE THE CAUSE OF BAMFORTH-LAZARUS
CC SYNDROME. A DISEASE ASSOCIATED WITH THYROID AGENESIS, CLEFT PALATE

```

CC AND CHOMAL ATRESTA.
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U89995; AAC51294.1; -;
 CC EMBL: Y13386; CAA73816.1; -;
 CC HSSP: Q63245; 2HFH.
 CC TRANSPAC: T02782; -;
 CC Genew: HGNC:3806; FOXE1.
 CC MIM: 602617; -;
 CC MIM: 241850; -;
 CC InterPro: IPR001766; TF_Fork_head.
 CC Pfam: PF00250; Fork_head; 1.
 CC PRINTS: PR00053; FORKHEAD.
 CC ProDom: PD000425; TF_Fork_head; 1.
 CC SMART: SM00339; FH; 1.
 CC PROSITE: PS00657; FORK_HEAD_1; 1.
 CC PROSITE: PS00658; FORK_HEAD_2; 1.
 CC PROSITE: PS50039; FORK_HEAD_3; 1.
 CC DNA-binding; Nuclear protein; Transcription regulation;
 CC Disease mutation.
 CC FT DOMAIN 7 12 POLY-PRO.
 CC FT DNA_BIND 52 143 FORK-HEAD.
 CC FT DOMAIN 164 182 POLY-ALA.
 CC FT VARIANT 65 65 A->V (IN BAMFORTH-LAZARUS SYNDROME).
 CC FT /FTID=VAR_008857.
 CC FT CONFLICT 178 182 MISSING (IN REF. 2).
 CC SQ SEQUENCE 376 AA; 38289 MM; 766534A73231572C C6C64;
 CC -----
 CC Query Match 9.1%; Score 98.5; DB 1; Length 376;
 CC Best Local Similarity 28.3%; Pred. No. 0.55;
 CC Matches 45; Conservative 13; Mismatches 68; Indels 33; Gaps 6;
 CC -----
 CC QY 12 NGWLTNPVGVAVSPDFYAGTVLLCQANQEGSSMYSGSSLYVTSAMGFPYPAATAAA 71
 CC Db 125 NYMALDPVAEDM---FSSGSFLRRKRKFSDDL-----STYPAYMHDAAAAAA 170
 CC QY 72 YRGALHRCGRGVNTFFAAAPPPPIPAYGVVQEPYGVNKLQGYAAVRYQPTPAT 131
 CC Db 171 AAAAAAIAAIFPGAVPAARPP---YPGAIV-----AGYAPPSLAPPPVY 215
 CC QY 132 AAAYSDSYGRVYA--ADPYHTTLAPAPTYGVGANNAPFA 167
 CC Db 216 YPAASPGPCRVFGLVPRPLSPGLPAPS-GRGSGCAFA 253
 CC -----
 CC RESULT 8
 CC EMS_MOUSE STANDARD; PRT; 655 AA.
 CC ID EMS_MOUSE
 CC AC 061545;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE RNA-binding protein Ems.
 CC GN Ems OR EWSH.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC OX NCBI_TaxID=10090;
 CC RP [1]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Testis;
 CC RX MEDLINE=95130099; PubMed=7829090;
 CC RA Ploungsael B., Mattei M.-G., Thomas G., Delattre O.,
 CC RT "Cloning and chromosome localization of the mouse Ews gene.",
 CC RL Genomics 23:278-281(1994).

CC -1- FUNCTION: MIGHT FUNCTION AS A REPRESSOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
 CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X79233; CAA5815.1; -;
 CC MGI: MGI:99960; Ewsb.
 CC DR InterPro: IPR000504; RNA_rec_mot.
 CC DR InterPro: IPR001876; Znf_RanBP.
 CC DR Pfam: PF00076; rrm; 1.
 CC DR Pfam: PF00641; zf-RanBP; 1.
 CC DR SMART: SM00360; RRM; 1.
 CC DR SMART: SM00547; Znf_RBP2; 1.
 CC DR PROSITE: PS50066; IQ; FALSE_NEG.
 CC DR PROSITE: PS50102; RRM; 1.
 CC DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 CC DR PROSITE: PS01358; ZF_RANBP2_1; 1.
 CC DR PROSITE: PS50199; ZF_RANBP2_2; 1.
 CC DR Nuclear protein; zinc-finger; zinc; RNA-binding; Metal-binding;
 CC KW Calmodulin-binding; Repeat; Methylation; Phosphorylation;
 CC KW Transcription regulation; Repressor.
 CC FT DOMAIN 1 285
 CC FT DOMAIN 256 485
 CC FT ZN_FING 360 446
 CC FT ZN_FING 517 548
 CC FT REPEAT 8 16
 CC FT REPEAT 17 27
 CC FT REPEAT 28 34
 CC FT REPEAT 35 42
 CC FT REPEAT 43 50
 CC FT REPEAT 51 59
 CC FT REPEAT 60 68
 CC FT REPEAT 69 75
 CC FT REPEAT 76 84
 CC FT REPEAT 85 91
 CC FT REPEAT 92 110
 CC FT REPEAT 111 116
 CC FT REPEAT 117 125
 CC FT REPEAT 126 156
 CC FT REPEAT 157 163
 CC FT REPEAT 164 170
 CC FT REPEAT 171 177
 CC FT REPEAT 178 188
 CC FT REPEAT 189 193
 CC FT REPEAT 194 201
 CC FT REPEAT 202 206
 CC FT REPEAT 207 212
 CC FT REPEAT 213 218
 CC FT REPEAT 219 224
 CC FT REPEAT 225 230
 CC FT REPEAT 231 238
 CC FT REPEAT 239 245
 CC FT REPEAT 246 252
 CC FT REPEAT 253 259
 CC FT REPEAT 260 276
 CC FT REPEAT 277 285
 CC FT DOMAIN 300 339
 CC FT DOMAIN 453 512
 CC FT DOMAIN 558 639
 CC FT MOD_RES 300 300
 CC FT MOD_RES 302 302
 CC FT MOD_RES 304 304
 CC -----
 CC IQ.
 CC RNA-BINDING (RRM).
 CC RANBP2-TYPE.
 CC 31 X APPROXIMATE TANDEM REPEATS.

```

FT MOD_RES 309 309 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 314 314 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 317 317 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 321 321 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 454 454 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 463 463 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 470 470 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 489 489 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 493 493 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 499 499 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 502 502 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 505 505 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 562 562 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 564 564 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 571 571 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 574 574 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 580 580 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 588 588 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 591 591 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 595 595 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 599 599 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 602 602 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 606 606 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 614 614 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 632 632 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 635 635 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 266 266 PHOSPHORYLATION (BY PKC) (BY SIMILARITY)
SQ SEQUENCE 655 AA; 68418 MW; 50735EB54247D69 CRC64;

```

```

Query Match 9.1%; Score 98.5; DB 1; Length 655;
Best Local Similarity 28.5%; Pred. No. 1;
Matches 49; Conservative 11; Mismatches 71; Indels 41; Gaps 9;

```

```

Qy 29 VAGTLLTQCA-NOEGSSWYSGPSLVTYAMGEPFPAATAAAYGALRGRGVNT 87
Db 29 YAGTT--OAYGQOSGYTGQPTDVSYTQQTATYTGQATVATSY-----GQPTGYST 79
Qy 88 FFAAAPPPPIPAVGVVYQEPVYGNKLLQGVAAVYVAPPTAATAAAYSDSYGVYAADP 147
Db 80 -----PFAPO-----AYGQPVGY-----GTGAYVSTTATVTTTQA-----SYAQS 116
Qy 148 YHHTLAPATYGVGMNAPRLDATRSHADVGLVLSLQASITQGYNR 199
Db 117 AYGTPOAPVPTYGQ-----QPTATAPTRPDGNGKPAETSPQSS--TGGYNQ 160

```

RESULT 9

```

HMEV_DROME STANDARD; PRT; 376 AA.
AC P06602; P07667; Q9V5B6;
DT 01-JAN-1988 (Rel. 06, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Segmentation protein even-skipped.
GN EVE OR CG3328.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_taxid=7227;
OK NCBI_taxid=7227;
RP SEQUENCE FROM N.A.
RX MEDLINE=87051744; PubMed=2877745;
RA McDonald P.M., Ingham P., Struhl G.;
RT "Isolation, structure, and expression of even-skipped: a second pair-
RL rule gene of Drosophila containing a homeo box.";
RL Cell 47:721-734(1986).
RP SEQUENCE FROM N.A.
RX MEDLINE=87218536; PubMed=2884106;
RA Fraech M., Hoey T., Rushlow C., Doyle H., Levine M.;
RT "Characterization and localization of the even-skipped protein of

```

```

RT Drosophila.";
RL EMBO J. 6:749-759(1987).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Goez C.J.,
RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Gale R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chapple M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKlos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Baer A., Baxterdale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Brinkman D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Chew S., Dahlke C., Davoport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kemnison J.A., Ketchum K.A.,
RA Jaisli B., Kalush F., Karpen G.H., Ke Z., Kemnison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Liaw P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclet J.M.,
RA Palazolo M., Plittman G.S., Pan S., Pollard J., Puri V., Shen M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach U.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RP [4]
RP SEQUENCE OF 1-58 FROM N.A.
RC STRAIN=OR-RC, WA-F, ZMS6, AF-S, and FL-S;
RX MEDLINE=96036621; PubMed=8524036;
RA Ludwig M.Z., Kreitman M.;
RT "Evolutionary dynamics of the enhancer region of even-skipped in
RL Drosophila.";
RL Mol. Biol. Evol. 12:1002-1011(1995).
RP [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 70-189.
RX MEDLINE=96134926; PubMed=8557047;
RA Hirsch J.A., Aggarwal A.K.;
RT "Structure of the even-skipped homeodomain complexed to AT-rich DNA:
RL new perspectives on homeodomain specificity.";
RL EMBO J. 14:6280-6291(1995).
RP [6]
RP FUNCTION: MAY PLAY A ROLE IN DETERMINING NEURONAL IDENTITY. MAY BE
CC DIRECTLY INVOLVED IN SPECIFYING IDENTITY OF INDIVIDUAL NEURONS.
CC PAIR-RULE PROTEIN REQUIRED FOR SEGMENTATION; INVOLVED IN
CC TRANSFORMING THE BROAD, SPATIAL, PERIODIC EXPRESSION PATTERNS OF
CC THE GAP GENES INTO A SYSTEM OF PRECISE PERIODIC EXPRESSION
CC PATTERNS OF THE PAIR-RULE AND SEGMENTARY POLARITY GENES.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEOBOX FAMILY.
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


Db 303 NSHDPHSEDTPTSDLEOPAKOFKORIKLGTQADVGLALGTLYGNVF 352

RESULT 11

EYA_DROME STANDARD: PRT: 766 AA.

AC 005201; Q9VMCL1; 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Developmental protein eyes absent (Protein Clift).

CLC OR EYA OR CG9554.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

CC [1]

CC SEQUENCE FROM N.A.

CC MEDLINE=93161413; PubMed=8431945;

CC Bontin N.M., Leliseon W.M., Benzer S.;

CC "The eyes absent gene: genetic control of cell survival and

CC differentiation in the developing Drosophila eye.";

CC Cell 72:379-395 (1993).

CC [2]

CC SEQUENCE FROM N.A.

CC STRAIN=Berkeley;

CC MEDLINE=20196006; PubMed=10731132;

CC Adame M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

CC Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

CC George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

CC Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

CC Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champs C.R., Pfeiffer B.D.,

CC Wan K.H., Doyle C., Baker E.G., Helt G., Nelson C.R., Miklos G.L.G.,

CC Abell J.F., Abmayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

CC Ballew R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,

CC Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,

CC Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,

CC Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

CC Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

CC de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

CC Dodson K., Doup L.E., Downe M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

CC Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

CC Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasseer K.,

CC Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

CC Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

CC Hoskins D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,

CC Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

CC Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

CC Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

CC Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

CC Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,

CC Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

CC Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,

CC Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,

CC Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

CC Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

CC Spleer E., Spredling A.C., Stapleton M., Strong R., Sun E.,

CC Svetske R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,

CC Wang Z.-Y., Wasserman D.A., Weislock G.M., Weisenbach J.,

CC Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

CC Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

CC Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

CC Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

CC "The genome sequence of Drosophila melanogaster.";

CC Science 287:2185-2195 (2000).

CC -1- FUNCTION: REQUIRED AT AN EARLY STAGE IN DEVELOPMENT OF THE

CC D. MELANOGASTER COMPOUND EYE.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: BELONGS TO THE EYA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

CC -----

CC EMBL, L08501; AAA28723.1; -

CC DR EMBL, L08502; AAA28110.1; -

CC DR EMBL, AE003614; AAF52400.1; -

CC DR FLYBase; FBgn0000320; c11.

CC DR InterPro; IPR001454; Hlgase/hydrase.

CC Pfam; PF00702; Hydrolase, 1.

CC Developmental protein; Nuclear protein; Alternative splicing.

CC KW DOMAIN 46 57 POLY-GLN.

CC FT DOMAIN 60 68 POLY-GLN.

CC FT DOMAIN 92 108 POLY-GLY.

CC FT DOMAIN 253 260 POLY-ALA.

CC FT DOMAIN 268 271 POLY-TYR.

CC FT DOMAIN 305 311 POLY-ALA.

CC FT DOMAIN 428 434 POLY-ALA.

CC FT VARSPLIC 1 23 MVLTPYVYAARCGLLIDKMTLP -> MLVNVPCYQNFSTL

CC FT DYY (IN ISOFORM 2).

CC SQ SEQUENCE 766 AA; 80656 MW; 3581C26AB1811E74 CRC64;

Query Match 8.8%; Score 96; DB 1; Length 766;

Best Local Similarity 23.7%; Pred. No. 1.9;

Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6;

QY 42 GSSMT-----SGPSLYVTSMPGPYPATRAAAYRGALRGRTVYTFRAAA 92

Db 229 GSNLGGSSASNPDLGVAANVSSAV-----AAAAAYVDK-----DYIYNSMOQYT 278

QY 93 PPPPIAYGVVYQRPVYGNKLLGCGYAAYRYAQPPTAA----- 134

Db 279 PPPFVSGVCTPRAAATTAARQAQMEGAAAAYLTPTAAAGNNNSQLVSPVAGYNNF 338

QY 135 -----YSDYSGRYVADPPYHHTLAP-----TYGCG-AMNAPALPTDATTBSHA 178

Db 339 GQDYGYYNEQYGNVSPANYPVAVSPSSASHGCHFHVAASSNLSSEPTDHS 395

RESULT 12

HDAD MOUSE STANDARD: PRT: 386 AA.

ID 062424;

AC 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-UN-2002 (Rel. 41, Last annotation update)

DE Homeobox protein Hox-A13 (Hox-1.10).

GN HOXA13 OR HOX-1.10.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

CC [1]

CC SEQUENCE FROM N.A.

CC MEDLINE=96259555; PubMed=8673126;

CC Mortlock D.P., Post L.C., Innis J.W.;

CC "The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads

CC to arrest of digital arch formation.";

CC Nat. Genet. 13:284-289 (1996).

CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY

CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A

CC CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH

CC STRUCTURES.

CC -1- SIMILARITY: BELONGS TO THE HOX-B HOMEBOX FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U59322; AAB03322.1; -;
DR HSSP; P14653; 1B72.
DR TRANSFAC; T03337; -;
DR MGD; MGI:96173; Hoxa13.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PSS0071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA BIND 38 51 POLY-ALA.
FT DOMAIN 320 379 HOMEBOX.
FT DOMAIN 52 57 POLY-GLY.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 73 84 POLY-ALA.
FT DOMAIN 101 104 POLY-ALA.
FT DOMAIN 116 133 POLY-ALA.
FT DOMAIN 198 205 POLY-ALA.
SQ SEQUENCE 386 AA; 39566 MW; 2B01DCC9B1951324 CRC64;

Query Match 8.7%; Score 94; DB 1; Length 386;
Best Local Similarity 23.9%; Pred. No. 1.3;
Matches 56; Conservative 15; Mismatches 91; Indels 72; Gaps 9;

QY 21 GAVYSPDFYACTVLLCOANGSGSMYS-----GPSLVYTSAMPGPFPYPAATAAAAYRG 76
DB 67 GGNSTVAAAAAANQCRNLMHAPLAPGAAAAAYSSAPGAPPSAAAAA 125
QY 77 LRGGRFTVYNTFRAPAPPPPIPAYGVVYQ-----EPYVGNLLGGYAAVRY-- 124
DB 126 -----AAAAAASGSGPAPGAPGAEAKQKSPCSAAQSSSGPALPYGYFGSGTYICA 182
QY 125 -----AQTPTAAAYSDS-----YGRVYAADPYNHT 151
DB 183 RMGHPNAIKSCAPASAAAFADKVMDTAGPAAEFFSSAKKEFAFYHGGYAAAGPYHHN 241
QY 152 LAAPPTY-----GVGAMNAFAPLTDKTRSHADVGLVLSLQASLYQGGIN 198
DB 242 -QPVPGYLDMPVVPGLGGPG-----ESRHEPLGLPMESYQPMALPNGWN 284

RESULT 13
HXAD_HUMAN STANDARD; PRT; 388 AA.
ID AC P31271; O43371;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A13 (Hox-13).
GN HOXA13 OR HOX13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=91712976; PubMed=9020844;
RT "Mutation of HOXA13 in hand-foot-genital syndrome.";
RL Nat. Genet. 15:179-180(1997).
RA Bradshaw H., Hinds K., Keppler D.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [1]
RP SEQUENCE OF 322-387 FROM N.A.
RX MEDLINE=9009876; PubMed=2574852;
RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT "The human HOX gene family.";
RL Nucleic Acids Res. 17:10385-10402(1989).
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.

DR EMBL; U82827; AAC50993.1; -;
DR EMBL; AC004080; -; NOT_ANNOTATED_CDS.
DR PIR; S14932; S14932.
DR HSSP; P14653; 1B72.
DR TRANSFAC; T03321; -;
DR Genew; HGNC:5102; HOXA13.
DR MIM; 142959; -;
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PSS0071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA BIND 322 381 HOMEBOX.
FT DOMAIN 38 53 POLY-ALA.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 73 84 POLY-ALA.
FT DOMAIN 116 133 POLY-ALA.
FT DOMAIN 145 150 POLY-ALA.
FT DOMAIN 200 207 POLY-ALA.
FT CONFLICT 146 146 A -> G (IN REF. 2).
FT CONFLICT 187 187 P -> H (IN REF. 2).
FT CONFLICT 195 195 P -> A (IN REF. 2).
FT CONFLICT 198 198 P -> A (IN REF. 2).
SQ SEQUENCE 388 AA; 39752 MW; 6CD9C9A561C2FF6 CRC64;

Query Match 8.7%; Score 94; DB 1; Length 388;
Best Local Similarity 23.6%; Pred. No. 1.3;
Matches 56; Conservative 18; Mismatches 87; Indels 76; Gaps 10;

QY 21 GAVYSPDFYACTVLLCOANGSGSMYSPSL-----VYTSAMPGPFPYPAATAAAAYRG 74
DB 67 GGNSTVAAAAAANQCRNLMHAPLAPGAAVSSA-PGAPPSAAAAA 124
QY 75 AHLRGRTVYNTFRAPAPPPPIPAYGVVYQ-----EPYVGNLLGGYAAVRY-- 124
DB 125 A-----AAAAAASGSGPAPGAPAAEAQKSPCSAAQSSSGPALPYGYFGSGTYIP 180
QY 125 -----AQTPTAAAYSDS-----YGRVYAADPY 148
DB 181 CARMGPPNAIKSCPPPSAAAAAFADKVMDTAGPAAEFFSSAKKEFAFYHGGYAAAGPY 240
QY 149 HTLAPAPTY-----GVGAMNAFAPLTDKTRSHADVGLVLSLQASLYQGGIN 198
DB 241 HHN-QPVPGYLDMPVVPGLGGPG-----ESRHEPLGLPMESYQPMALPNGWN 286

RESULT 14
CH36_CERCA

```

ID CH36_CERCA STANDARD; PRT; 320 AA.
AC P17110;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chorion protein S36.
GN CP36 OR S36.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Tephriticoidea; Tephritidae; Ceratitis.
OC NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245561; PubMed=1692403;
RA Konojaki M., Konitopoulou K., Tollas P.P., King D.L., Swimmer C.,
RA Kataros F.C.;
RT "The chorion genes of the medfly, Ceratitis capitata, I: Structural
RT and regulatory conservation of the s36 gene relative to two
RT Drosophila species."
RL Nucleic Acids Res. 18:1731-1737(1990).
CC -I- DOMAIN: THE TETRAPEPTIDE (A-A-P-[AV]) REPEATS FOUND THROUGHOUT THE
CC PROTEIN ARE ALSO PRESENT IN MANY PROTEINS CONSTITUTING THE
CC PROTECTIVE ENVELOPE OF OTHER SPECIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51342; CAA35723.1; -.
DR PIR: S09208; S09208.
DR Chorion; Repeat.
FT REPEAT 178 181 1.
FT REPEAT 258 261 2.
FT REPEAT 266 269 3.
FT REPEAT 274 277 4.
FT REPEAT 280 290 5.
SQ SEQUENCE 320 AA; 32319 MW; EC9B72FFB854C61 CRC64;

Query Match 8.6%; Score 93.5; DB 1; Length 320;
Best Local Similarity 24.7%; Pred. No. 1.2;
Matches 40; Conservative 11; Mismatches 48; Indels 63; Gaps 6;

QY 4 KKAANDPYTNGMKLNPNVAGVYSP-----DFYAGTVLLQANDGSSMTSGPSSLVYTS4 57
DB 193 QGVINKVPTPLSLNPNVYVYKPKKIDAPLVGVQONVQAPSGSSYSAP----- 244
QY 58 MGFPRPATTAATAAAYGAILRGGRVYNTFRRAAPRPPIPAVGGVYQEPVYGNKLLQ8 117
DB 245 -----AASTY-----PAPPSYSAAPAO----- 262
QY 118 GYAAAYRAOPTATAAAYSDSYGRVYADPYHHTLAPAPTYG 159
DB 263 ----SYN-AAPAPSYSAAPASYSY-----AAPSASYDAPAPASYSY 297

RESULT 15
GAT6_CHICK STANDARD; PRT; 387 AA.
AC P43563;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor GATA-6 (GATA binding factor-6).
GN GATA6.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

```

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=white leghorn;
RX MEDLINE=94365018; PubMed=8083222;
RA Laverriere A.C., Macneill C., Mueller C., Poelmann R.E.,
RA Butch J.B.E., Evans T.;
RT "GATA-4/5/6, a subfamily of three transcription factors transcribed
RT in developing heart and gut."
RT J. Biol. Chem. 269:23177-23184(1994).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH, AND IN SMALL
CC INTESTINE. LOWER LEVELS IN LUNG, LIVER, OVARY AND HEART.
CC -----
CC -I- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U1889; AAA57505.1; -.
DR HSPB; P17679; IGNF.
DR TRANSFAC; T02684; -.
DR InterPro; IPR000679; Znf_GATA.
DR InterPro; IPR001164; hrip_like.
DR Pfam; PF00320; GATA; 2.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 2.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 2.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT ZN_FING 181 205 GATA-TYPE 1.
FT ZN_FING 235 259 GATA-TYPE 2.
FT DOMAIN 240 244 POLY-THR.
SQ SEQUENCE 387 AA; 40249 MW; 6EEF457BB34DB2C0 CRC64;

Query Match 8.6%; Score 93; DB 1; Length 387;
Best Local Similarity 26.0%; Pred. No. 1.6;
Matches 47; Conservative 15; Mismatches 51; Indels 68; Gaps 10;

QY 20 VGAVYSPPYACTVLLQANDGSSMTSGPSSLVYT-----SAMPGFY----- 63
DB 7 IASQGPARYDGS-----PGGFHSAFSSPYVPTTRVGSVLPYLYGGGAAQPG 58
QY 64 -----PATATAA-AYGAHLRGGRVYNTFRRAAPRPPI-----PAYGVVYQ 106
DB 59 HAPAGHWSQPAAESPSYGAAGAPSGR-----FPYSASPPVANGASREQYGGGLAA 111
QY 107 EPVYG--NKLLGGYAAAYRAOPTATAAAY--SDSYGRVYADPYHHTL-----APA 155
DB 112 REQYGLPRLPLNGSY-----PAPYASYGPOLGPRMPAPFENSVLHLGQRAAI 162
QY 156 P 156
DB 163 P 163

Search completed: March 12, 2003, 22:24:41
Job time : 19 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:22:11 ; Search time 43 Seconds
(without alignments)
972.734 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MTKKAVNPYNGMKLNPV.....VLSSQASIVQGGYRFPAY 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTRMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	1063	97.9	418	4	Q8WYB2
2	1061	97.7	417	11	Q8R4Z7
3	912	84.0	392	4	Q8TAE3
4	729.5	67.2	395	4	Q8TAE3
5	676	62.2	376	6	Q95K10
6	532.5	49.0	293	4	Q9UGW3
7	526	48.4	380	4	Q8WYB1
8	526	48.4	390	4	Q8TDD0
9	525	48.3	377	11	Q923W8
10	525	48.3	377	11	Q8V162
11	522	48.1	358	4	Q96DZ6
12	519	47.8	450	4	Q96NL7
13	456	42.0	303	11	Q8V161
14	351	32.3	263	4	Q8TCH0
15	296	27.3	435	11	Q8R2T5
16	129	11.9	367	5	Q9VT99

17	123.5	11.4	326	5	Q9NFX5	Q9NFX5 ceratilis c
18	116.5	10.7	326	5	Q62010	Q62010 ceratilis c
19	110.5	10.2	219	5	Q9V875	Q9V875 drosophila
20	108.5	10.0	291	4	Q92567	Q92567 homo sapien
21	104.5	9.6	226	5	Q95V16	Q95V16 myzus persi
22	103	9.5	332	5	Q9NFX7	Q9NFX7 ceratilis c
23	103	9.5	338	16	Q9R233	Q9R233 delnoccocus
24	102	9.4	604	16	Q8XPW5	Q8XPW5 ralsotonia s
25	101.5	9.3	950	5	Q9VE45	Q9VE45 drosophila
26	100	9.2	830	5	Q9W176	Q9W176 drosophila
27	99	9.1	542	5	Q23446	Q23446 caenorhabdi
28	98.5	9.1	723	11	Q9YK84	Q9YK84 mus musculu
29	98	9.0	316	5	Q16463	Q16463 caenorhabdi
30	97.5	9.0	332	5	Q62009	Q62009 ceratilis c
31	97.5	9.0	358	10	Q94223	Q94223 oryza sativ
32	97.5	9.0	828	5	Q9M349	Q9M349 drosophila
33	97	8.9	179	5	Q9V118	Q9V118 drosophila
34	97	8.9	363	5	Q01353	Q01353 drosophila
35	97	8.9	584	16	Q9FCJ3	Q9FCJ3 streptomyce
36	97	8.9	723	11	Q9WVH2	Q9WVH2 mus musculu
37	97	8.9	811	16	Q9X8T0	Q9X8T0 streptomyce
38	97	8.9	835	5	Q95YR9	Q95YR9 leishmania
39	96.5	8.9	131	10	Q40415	Q40415 nicotiana s
40	96.5	8.9	748	10	Q9T0G5	Q9T0G5 arabidopsis
41	96.5	8.9	1024	12	Q89815	Q89815 bovine herp
42	96	8.8	760	5	Q961V4	Q961V4 drosophila
43	95.5	8.8	200	5	P82167	P82167 locusta mig
44	95.5	8.8	392	5	Q9BLS3	Q9BLS3 leishmania
45	95	8.7	275	5	Q9BPN1	Q9BPN1 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q8WYB2	PRELIMINARY;	PRT;	418 AA.
AC	Q8WYB2;			
DT	01-MAR-2002 (TREMURel. 20, Created)			
DT	01-MAR-2002 (TREMURel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMURel. 21, Last annotation update)			
DE	Hexaribonucleotide binding protein 1 isoform gamma.			
GN	HRNBPL.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Chen W., Winkelman J.C.;			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelman J.C.;			
RT	"Molecular Cloning and Chromosomal Localization of a Novel Human			
RT	Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene			
RT	Homologous to fox-1 in Caenorhabditis elegans."			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Chen W., Winkelman J.C.;			
RL	"The Exon-Intron Organization of the Human HRNBPL Gene."			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF229057; AAL71904.1; -			
DR	EMBL; AF109120; AAL83409.1; -			
DR	EMBL; AF109107; AAL83409.1; JOINED.			
DR	EMBL; AF109108; AAL83409.1; JOINED.			
DR	EMBL; AF109109; AAL83409.1; JOINED.			
DR	EMBL; AF109110; AAL83409.1; JOINED.			
DR	EMBL; AF109111; AAL83409.1; JOINED.			
DR	EMBL; AF109112; AAL83409.1; JOINED.			
DR	EMBL; AF109113; AAL83409.1; JOINED.			
DR	EMBL; AF109114; AAL83409.1; JOINED.			


```

AC 08TAF2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 1 isoform beta.
GN HRNBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RT "The Exon-Intron Organization of the Human HRNBP1 Gene.";
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109106; AAL83406.1; -.
DR EMBL; AF109120; AAL83407.1; JOINED.
DR EMBL; AF109107; AAL83407.1; JOINED.
DR EMBL; AF109108; AAL83407.1; JOINED.
DR EMBL; AF109109; AAL83407.1; JOINED.
DR EMBL; AF109110; AAL83407.1; JOINED.
DR EMBL; AF109111; AAL83407.1; JOINED.
DR EMBL; AF109112; AAL83407.1; JOINED.
DR EMBL; AF109113; AAL83407.1; JOINED.
DR EMBL; AF109114; AAL83407.1; JOINED.
DR EMBL; AF109115; AAL83407.1; JOINED.
DR EMBL; AF109116; AAL83407.1; JOINED.
DR EMBL; AF109117; AAL83407.1; JOINED.
DR EMBL; AF109118; AAL83407.1; JOINED.
DR EMBL; AF109119; AAL83407.1; JOINED.
SQ SEQUENCE 395 AA; 42402 MW; 2299F9127A1DE39 CRC64;

Query Match 67.2%; Score 729.5; DB 4; Length 395;
Best Local Similarity 81.2%; Pred. No. 3.2e-52;
Matches 143; Conservative 2; Mismatches 12; Indels 19; Gaps 2;

QY 1 MTNKAANVPYNGMKLNPNVGAAYSPDPFAGTLLCQANQSGSSMYSGPSSLVYTSAMPG 60
DB 216 MTNKTVPYNGMKLNPNVGAAYSPDPFAGTLLCQANQSGSSMYSGPSSLVYTSAMPG 275
QY 61 PPYPATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGVVYQEPYGNKLLQGGYA 120
DB 276 PPYPATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGVVYQEPYGNKLLQGGYA 335
QY 121 AYRYAQPTPATAAAYSDSYGRVY-AADEPH-----HTLAPAPT 157
DB 336 AYRYAQPTPATAAAYSDSYGRVY-AADEPH-----HTLAPAPT 391

RESULT 5
Q95K10 PRELIMINARY; PRT; 376 AA.
AC Q95K10;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.6 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissus-TEMPORAL LOBE RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,

```

```

RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB060859; BAB46877.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 40565 MW; DAA7A4DD07B1030A CRC64;

Query Match 62.2%; Score 676; DB 6; Length 376;
Best Local Similarity 76.7%; Pred. No. 7.6e-48;
Matches 135; Conservative 4; Mismatches 17; Indels 20; Gaps 3;

QY 1 MTNKAANVPYNGMKLNPNVGAAYSPDPFAGTLLCQANQSGSSMYSGPSSLVYTSAMPG 60
DB 198 MTNKTVPYNGMKLNPNVGAAYSPDPFAGTLLCQANQSGSSMYSGPSSLVYTSAMPG 257
QY 61 PPYPATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGVVYQEPYGNKLLQGGYA 120
DB 258 PPYPATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGVVYQEPYGNKLLQGGYA 316
QY 121 AYRYAQPTPATAAAYSDSYGRVY-AADEPH-----HTLAPAPT 157
DB 317 AYRYAQPTPATAAAYSDSYGRVY-AADEPH-----HTLAPAPT 372

RESULT 6
Q9UGW3 PRELIMINARY; PRT; 293 AA.
AC Q9UGW3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE D41P2.2 (Supported by GENSCAN) (Fragment).
GN RBM9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049748; CAB3055.1; -.
DR HSSP; P11940; ICVJ.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
FT NON TER 1
SQ SEQUENCE 293 AA; 30983 MW; F9A823DD70CC1004 CRC64;

Query Match 49.0%; Score 532.5; DB 4; Length 293;
Best Local Similarity 52.8%; Pred. No. 3.6e-36;
Matches 114; Conservative 17; Mismatches 44; Indels 41; Gaps 7;

QY 1 MTNKAANVPYNGMKLNPNVGAAYSPDPFAGTLLCQANQSGSSMYSGPSSLVYTSAMPG 56
DB 106 MTNKTVPYNGMKLNPNVGAAYSPDPFAGTLLCQANQSGSSMYSGPSSLVYTSAMPG 165
QY 57 AMPGPPY-AAATAAAYGAHLRGRGRVTYNTFRAAAPPPPIPAYGVVYQEPYGNKLL 115
DB 166 IIPGPPYTAATAAAYGAHLRGRGRVTYNTFRAAAPPPPIPAYGVVYQEPYGNKLL 224
QY 116 OGGYAAAYGAQPTPAT-----AAAYSDSYGRVYAADPVHHTLAPAPTYGAMNFA 167
DB 225 OGGYAAAYGAQPTPATATAAATAAAYSDSYGRVYAADPVHHTLAPAPTYGAMNFA 279
QY 168 PLTDAKTRSHADVGLVLSLQASLYOGYNRFAPY 203

```

Db 280 -----ASLYRGYSRFAPY 293

RESULT 7

Q8WYB1 PRELIMINARY; PRT; 380 AA.

AC Q8WYB1; 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hexaribonucleotide binding protein 2 (RNA binding motif protein 9).
 GN HRNP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Chen W., Winkelmann J.C.;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE=UTERUS;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC025281; AAL71905.1; -
 DR InterPro; IPR000104; Antifreeze_1.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR PRINTS; PRO0308; ANTI-FREEZE.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 SQ SEQUENCE 380 AA; 40351 MW; 601P4E31E7C20715 CRC64;

Query Match 48.4%; Score 526; DB 4; Length 380;
 Best Local Similarity 52.7%; Pred. No. 1.7e-35;
 Matches 116; Conservative 15; Mismatches 43; Indels 46; Gaps 8;

QY 1 MTNKAANPYNTGKMLNPVGAIVSPDYAGTVLLCOAN--QEGSSMVGSPSS-----L 52
 DB 190 MTNKKVNTFYANGMKLSPVVGAVYGPBELYAASSFOADVSLGNDAAVPLSGRGGINTYIPL 249
 QY 53 VYTSAMPGEPPY-AAATAAAYRGALHRCGRVTYNTFRFAAPPPPIPAVGVVYQEPVYG 111
 DB 250 ISLPLVPGFPYPTAATTAATAFRGALHRCGRVTYGAVR-AVPTAIAPVGVVYQDGFYG 308
 QY 112 NKLLOGVAAYRYAQPPTPAT-----AAAYSDSYGRVYAADPYHHTLAPAPTYGVGAM 163
 DB 309 AD-LVGGYAAYRYAQPATTAATAAAAAAAYSDYGRVYTADPY-HALAPAASTYGVGAV 366
 QY 164 NAFAPLTDKTRSHADVGLVSLQASLYOGGYNRFAPY 203
 DB 367 -----ASLYRGYSRFAPY 380

RESULT 8

Q8TD00 PRELIMINARY; PRT; 390 AA.

AC Q8TD00; 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE RNA binding motif protein 9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.

RX MEDLINE=21864496; PubMed=11875103;
 RA Norris J.D., Fan D., Sheik A., McDonnell D.P.;
 RT "A Negative Coregulator for the Human ER."
 RL Mol. Endocrinol. 16:459-468(2002).
 DR EMBL; AY072786; AAL67150.1; -
 SQ SEQUENCE 390 AA; 41374 MW; 40D6C8C40764B317 CRC64;

Query Match 48.4%; Score 525; DB 11; Length 390;
 Best Local Similarity 53.7%; Pred. No. 2e-35;
 Matches 116; Conservative 14; Mismatches 44; Indels 42; Gaps 8;

QY 1 MTNKAANPYNTGKMLNPVGAIVSPDYAGTVLLCOAN--QEGSSMVGSPSS-----L 52
 DB 200 MTNKKVNTFYANGMKLSPVVGAVYGPBELYAASSFOADVSLGNDAAVPLSGRGGINTYIPL 259
 QY 53 VYTSAMPGEPPY-AAATAAAYRGALHRCGRVTYNTFRFAAPPPPIPAVGVVYQEPVYG 111
 DB 260 ISLPLVPGFPYPTAATTAATAFRGALHRCGRVTYGAVR-AVPTAIAPVGVVYQDGFYG 318
 QY 112 NKLLOGVAAYRYAQPPTPAT-----AAAYSDSYGRVYAADPYHHTLAPAPTYGVGAM 163
 DB 319 AD-LVGGYAAYRYAQPATTAATAAAAAAAYSDYGRVYTADPY-HALAPAASTYGVGAV 376
 QY 164 NAFAPLTDKTRSHADVGLVSLQASLYOGGYNRFAPY 203
 DB 377 -----ASLYRGYSRFAPY 390

RESULT 9

Q923W8 PRELIMINARY; PRT; 377 AA.

AC Q923W8; 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative RNA-binding protein fxh.
 GN FXH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21294761; PubMed=11401487;
 RA Lieberman A.P., Friedlich D.L., Harmison G., Howell B.W., Jordan C.L., Breedlove S.M., Fischbeck K.H.;
 RT "Androgens Regulate the Mammalian Homologues of Invertebrate Sex Determination Genes tra-2 and fox-1."
 RL Biochem. Biophys. Res. Commun. 282:499-506(2001).
 DR EMBL; AF387322; AAK64287.1; -
 DR MGD; MGI:1933973; Fxh.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 SQ SEQUENCE 377 AA; 40165 MW; 249FA8A6A0D155A CRC64;

Query Match 48.3%; Score 525; DB 11; Length 377;
 Best Local Similarity 53.7%; Pred. No. 2e-35;
 Matches 116; Conservative 14; Mismatches 44; Indels 42; Gaps 8;

QY 1 MTNKAANPYNTGKMLNPVGAIVSPDYAGTVLLCOAN--QEGSSMVGSPSS--VYTS 56
 DB 191 MTNKKVNTFYANGMKLSPVVGAVYGPBELYAASSFOADVSLGNDAAVPLSGRGGINTYIPL 250
 QY 57 AMPGFPY-AAATAAAYRGALHRCGRVTYNTFRFAAPPPPIPAVGVVYQEPVYGKLL 115
 DB 251 IIFGFPYPTAATTAATAFRGALHRCGRVTYGAVR-AVPTAIAPVGVVYQDGFYAD-L 308
 QY 116 OGGYAAYRYAQPPTPAT-----AAAYSDSYGRVYAADPYHHTLAPAPTYGVGAMNAPA 167
 DB 309 YGGYAAYRYAQPATTAATAAAAAAAYSDYGRVYTADPY-HALAPAASTYGVGAV--- 363

Qy 168 PLTDKTRSHADVGLVLSLQASLYOGGYNRFAPY 203
Db 364 -----ASLYRGYSRFAPY 377

RESULT 10

ID Q9VI62 PRELIMINARY; PRT; 377 AA.
AC Q9VI62;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hexabinucleotide binding protein 2.
GN FXH OR HRNB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229055; AAL71902.1; -.
DR MGD; MGI:1933973; Fxh.
DR InterPro; IPR00104; Antifreeze_1.
DR Pfam; PF00076; rrm; 1.
DR PRINTS; PRO0308; ANTIFREEZE1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM RNP 1; UNKNOWN 1.
SQ SEQUENCE 377 AA; 40183 MW; 249E72CAFA7C505A CRC64;

Query Match 48.3%; Score 525; DB 11; Length 377;
Best Local Similarity 53.7%; Pred. No. 2e-35; Matches 115; Conservative 14; Mismatches 44; Indels 42; Gaps 8;

Qy 1 MTNKAANPYNGKLNPPVGA VSPDPYAGTVLLCOAN--QEGSSMWGSPSSL--VYTS 56
Db 191 MTNKAANPYNGKLNPPVGA VSPDPYAGTVLLCOAN--QEGSSMWGSPSSL--VYTS 56
Qy 57 AMGFPYP-AATAAAYRGALHGRGRTVNTFRAAPPPPIPAYGVVYQEPYGNKL 115
Db 251 IIPGFPYPTAATTAAGAHLRGGRGVYAVR-AVPTAIIPAYPGVVYQDGFYAD-L 308
Qy 116 QCGAAARYAOPTRAT-----AAASDSYGRVYADPYHTHTLAPRYGVGAMNAFA 167
Db 309 YGGAAARYAOPATATATAATAAAAAAASDYGRTYADPY-HALAPAAAYGVGAV---- 363
Qy 168 PLTDKTRSHADVGLVLSLQASLYOGGYNRFAPY 203
Db 364 -----ASLYRGYSRFAPY 377

RESULT 11

ID Q96DZ6 PRELIMINARY; PRT; 358 AA.
AC Q96DZ6;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Similar to fox-1 homolog (C. elegans).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Tissue:ADRENAL GLAND;
RA Straussberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013115; AAH13115.1; -.
DR InterPro; IPR00504; RNA_rec_mot.

DR Pfam; PF00076; rrm; 1.
DR PROSITE; PSS0102; RRM; 1.
SQ SEQUENCE 358 AA; 37907 MW; BA4B3740C178E441 CRC64;

Query Match 48.1%; Score 522; DB 4; Length 358;
Best Local Similarity 53.2%; Pred. No. 3.4e-35; Matches 115; Conservative 15; Mismatches 44; Indels 42; Gaps 8;

Qy 1 MTNKAANPYNGKLNPPVGA VSPDPYAGTVLLCOAN--QEGSSMWGSPSSL--VYTS 56
Db 172 MTNKAANPYNGKLNPPVGA VSPDPYAGTVLLCOAN--QEGSSMWGSPSSL--VYTS 56
Qy 57 AMGFPYP-AATAAAYRGALHGRGRTVNTFRAAPPPPIPAYGVVYQEPYGNKL 115
Db 232 IIPGFPYPTAATTAAGAHLRGGRGVYAVR-AVPTAIIPAYPGVVYQDGFYAD-L 289
Qy 116 QCGAAARYAOPTRAT-----AAASDSYGRVYADPYHTHTLAPRYGVGAMNAFA 167
Db 290 YGGAAARYAOPATATATAATAAAAAAASDYGRTYADPY-HALAPAAAYGVGAV---- 344
Qy 168 PLTDKTRSHADVGLVLSLQASLYOGGYNRFAPY 203
Db 345 -----ASLYRGYSRFAPY 358

RESULT 12

ID Q96NL7 PRELIMINARY; PRT; 450 AA.
AC Q96NL7;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE CDNA FLJ30651 fis, clone D30ST2000184, highly similar to RNA binding motif protein 9.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Nishimura K., Nagasawa M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Aoe K., Kamihara K.,
RA Otsuka N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Negahari K., Maehno Y., Negai K., Isogai T.,
RL "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055213; BAB70875.1; -.
DR InterPro; IPR00504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM RNP 1; UNKNOWN 1.
SQ SEQUENCE 450 AA; 47236 MW; 173D0C55D5CD762 CRC64;

Query Match 47.8%; Score 519; DB 4; Length 450;
Best Local Similarity 52.3%; Pred. No. 7.8e-35; Matches 115; Conservative 14; Mismatches 45; Indels 46; Gaps 8;

Qy 1 MTNKAANPYNGKLNPPVGA VSPDPYAGTVLLCOAN--QEGSSMWGSPSSL--L 52
Db 260 MTNKAANPYNGKLNPPVGA VSPDPYAGTVLLCOAN--QEGSSMWGSPSSL--L 52
Qy 53 VYTSAMGFPYP-AATAAAYRGALHGRGRTVNTFRAAPPPPIPAYGVVYQEPYGN 111
Db 320 ISLPLVPGFPYPTAATTAAGAHLRGGRGVYAVR-AVPTAIIPAYPGVVYQDGFYG 378
Qy 112 NKLOGGAAARYAOPTRAT-----AAASDSYGRVYADPYHTHTLAPRYGVGAM 163
Db 379 AD-LYGGAAARYAOPATATATAATAAAAAAASDYGRTYADPY-HALAPAAAYGVGAV 436

QY 164 NAFAPLTDKTRSHADDVGLVSLQASLYOGGYNRFAPY 203
 Db 437 -----ASLYRGYSRFRAPY 450

RESULT 13

ID Q8VI61 PRELIMINARY; PRT; 303 AA.
 AC Q8VI61;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hexaribonucleotide binding protein 3 (Fragment).
 GN HRNBP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RA Chen W., Minkelmann J.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF229056; AAL71903.1; -.
 DR InterPro; IPR005054; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 303 AA; 32864 MW; 74B0AF9F9FB7935 CRC64;

Query Match 42.0%; Score 456; DB 11; Length 303;
 Best Local Similarity 45.5%; Pred. No. 7.5e-30;
 Matches 95; Conservative 6; Mismatches 32; Indels 76; Gaps 4;

QY 3 NKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSSLVYTSAMPGP 62
 Db 123 NNATARVMTNGMKLNIPVGVGVGPEFYAVT-----SFP 155
 QY 63 YPAATAAAYRGAHLRGRRTVYNTFRAPAPPPIPAYGG----- 102
 Db 156 YTTGTAAYVRGTHLKGRAVYNTFRAPAPPPIPTTGALAEQTLVGNPVPAGLAPCP 215
 QY 103 -----VYQEFVYGNKLLGGYAAVRAQPTPATAAAY 135
 Db 216 LPQQTPEBAPYTPSPAFPPPLSCFPASRVVYQDGFYGA- IYGVYAAVRAQPAATPAAY 274
 QY 136 SDSYGRVY-AADPYHHTLAPAPTYGVGM 163
 Db 275 SDSYGRVYAADPYHHTIGTATYSIGTM 303

RESULT 14

ID Q8TCM0 PRELIMINARY; PRT; 263 AA.
 AC Q8TCM0;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 28.6 kDa protein (Fragment).
 GN DKFZP547L059.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRN;
 RA Blocker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL71700; CAD28499.1; -.
 KW Hypothetical protein.

FT NON_TER 263 263
 SQ SEQUENCE 263 AA; 28565 MW; 56C2D8538468CE1 CRC64;

Query Match 32.3%; Score 351; DB 4; Length 263;
 Best Local Similarity 35.6%; Pred. No. 2.8e-21;
 Matches 65; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTNKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSSLVYTSAMPGP 60
 Db 196 MTNKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOGSSMYSAFSSLVYTSAMPGP 255
 QY 61 FFPYATA 68
 Db 256 FFPYATA 263

RESULT 15

ID Q8R2T5 PRELIMINARY; PRT; 435 AA.
 AC Q8R2T5;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similar to RNA binding motif protein 9.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Straubeberg R.;
 RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC027263; AAH27263.1; -.
 SQ SEQUENCE 435 AA; 46252 MW; 62A355605FF8A5B4 CRC64;

Query Match 27.3%; Score 296; DB 11; Length 435;
 Best Local Similarity 50.4%; Pred. No. 1.7e-16;
 Matches 70; Conservative 10; Mismatches 53; Indels 6; Gaps 4;

QY 1 MTNKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOAN--OEGSSMYSGPSSL--VYTS 56
 Db 259 MTNKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOAN--OEGSSMYSGPSSL--VYTS 318
 QY 57 AMPGPPYP-AATPAAYRGAHLRGRRTVYNTFRAPAPPPIPAYGVYQEPVYGNKLL 115
 Db 319 IIPGPPYPYATTAAPRGAHLRGRRTVYGAVR-AVPEYALPAPGVDMQPTDMSLL 377
 QY 116 OGVYAAVRYAOPTPATAA 134
 Db 378 QPQPOLLPLOPLTATVTA 396

Search completed: March 13, 2003, 00:54:09
 Job time : 44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:24:21 ; Search time 19 Seconds
(without alignments)

314.361 Million cells updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MTNKKAVNPYTNQWKLNPV.....VLSLSQASIVQGGYRFPAPY 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents AA:*
2: /cgn2_6/ptodata/1/iaa/5A COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/5B COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6A COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/6B COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/6C COMB.pep:*
7: /cgn2_6/ptodata/1/iaa/6D COMB.pep:*
8: /cgn2_6/ptodata/1/iaa/6E COMB.pep:*
9: /cgn2_6/ptodata/1/iaa/6F COMB.pep:*
10: /cgn2_6/ptodata/1/iaa/6G COMB.pep:*
11: /cgn2_6/ptodata/1/iaa/6H COMB.pep:*
12: /cgn2_6/ptodata/1/iaa/6I COMB.pep:*
13: /cgn2_6/ptodata/1/iaa/6J COMB.pep:*
14: /cgn2_6/ptodata/1/iaa/6K COMB.pep:*
15: /cgn2_6/ptodata/1/iaa/6L COMB.pep:*
16: /cgn2_6/ptodata/1/iaa/6M COMB.pep:*
17: /cgn2_6/ptodata/1/iaa/6N COMB.pep:*
18: /cgn2_6/ptodata/1/iaa/6O COMB.pep:*
19: /cgn2_6/ptodata/1/iaa/6P COMB.pep:*
20: /cgn2_6/ptodata/1/iaa/6Q COMB.pep:*
21: /cgn2_6/ptodata/1/iaa/6R COMB.pep:*
22: /cgn2_6/ptodata/1/iaa/6S COMB.pep:*
23: /cgn2_6/ptodata/1/iaa/6T COMB.pep:*
24: /cgn2_6/ptodata/1/iaa/6U COMB.pep:*
25: /cgn2_6/ptodata/1/iaa/6V COMB.pep:*
26: /cgn2_6/ptodata/1/iaa/6W COMB.pep:*
27: /cgn2_6/ptodata/1/iaa/6X COMB.pep:*
28: /cgn2_6/ptodata/1/iaa/6Y COMB.pep:*
29: /cgn2_6/ptodata/1/iaa/6Z COMB.pep:*
30: /cgn2_6/ptodata/1/iaa/6A COMB.pep:*
31: /cgn2_6/ptodata/1/iaa/6B COMB.pep:*
32: /cgn2_6/ptodata/1/iaa/6C COMB.pep:*
33: /cgn2_6/ptodata/1/iaa/6D COMB.pep:*
34: /cgn2_6/ptodata/1/iaa/6E COMB.pep:*
35: /cgn2_6/ptodata/1/iaa/6F COMB.pep:*
36: /cgn2_6/ptodata/1/iaa/6G COMB.pep:*
37: /cgn2_6/ptodata/1/iaa/6H COMB.pep:*
38: /cgn2_6/ptodata/1/iaa/6I COMB.pep:*
39: /cgn2_6/ptodata/1/iaa/6J COMB.pep:*
40: /cgn2_6/ptodata/1/iaa/6K COMB.pep:*
41: /cgn2_6/ptodata/1/iaa/6L COMB.pep:*
42: /cgn2_6/ptodata/1/iaa/6M COMB.pep:*
43: /cgn2_6/ptodata/1/iaa/6N COMB.pep:*
44: /cgn2_6/ptodata/1/iaa/6O COMB.pep:*
45: /cgn2_6/ptodata/1/iaa/6P COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	551	50.7	330	4	US-09-145-391-2
2	288	26.5	366	3	US-09-176-657-2
3	96	8.8	760	1	US-08-195-152-2
4	94	8.7	349	4	US-09-343-011B-1
5	93	8.6	444	4	US-09-252-292C-27
6	92.5	8.5	362	1	US-08-437-027-21
7	92.5	8.5	365	1	US-08-437-027-20
8	92.5	8.5	656	2	US-08-343-443B-2
9	92.5	8.5	656	4	US-09-214-564A-4
10	91	8.4	346	4	US-09-343-011B-2
11	87.5	8.1	980	4	US-09-442-100-8
12	87	8.0	379	1	US-08-552-142A-11
13	86.5	8.0	262	3	US-08-946-914-14
14	86.5	8.0	262	4	US-08-656-450-14
15	86.5	8.0	4302	3	US-08-658-136-5
16	86.5	8.0	4302	4	US-09-052-469-8
17	86.5	8.0	4339	4	US-09-052-469-6
18	86	7.9	207	4	US-09-336-536-50
19	86	7.9	245	4	US-09-336-536-48
20	85	7.8	275	1	US-08-431-387-3
21	85	7.8	275	1	US-08-322-677A-7
22	85	7.8	275	1	US-08-322-676-7
23	85	7.8	275	1	US-08-460-343B-74
24	85	7.8	275	1	US-08-398-028B-74
25	85	7.8	275	1	US-08-504-265B-90
26	85	7.8	275	2	US-08-140-083A-9
27	85	7.8	275	2	US-08-865-203-8

28	85	7.8	275	2	US-09-135-658-3	Sequence 3, Appli
29	85	7.8	275	2	US-07-849-420-8	Sequence 8, Appli
30	85	7.8	275	3	US-08-898-218-7	Sequence 7, Appli
31	85	7.8	275	3	US-08-848-793-7	Sequence 7, Appli
32	85	7.8	275	4	US-09-253-854-8	Sequence 8, Appli
33	85	7.8	275	4	US-08-955-424-8	Sequence 8, Appli
34	85	7.8	275	4	US-09-178-155-3	Sequence 3, Appli
35	85	7.8	275	4	US-09-445-270-2	Sequence 2, Appli
36	85	7.8	275	4	US-09-467-536A-2	Sequence 2, Appli
37	85	7.8	275	4	US-09-234-957-2	Sequence 2, Appli
38	85	7.8	275	4	US-08-394-011-1	Sequence 1, Appli
39	85	7.8	275	4	US-08-397-329-1	Sequence 1, Appli
40	85	7.8	275	5	PCT-US93-00418-3	Sequence 1, Appli
41	85	7.8	275	5	PCT-US95-04760-2	Sequence 2, Appli
42	85	7.8	352	1	US-07-923-260A-2	Sequence 2, Appli
43	85	7.8	352	1	US-07-923-260A-17	Sequence 3, Appli
44	85	7.8	380	1	US-08-552-142A-17	Sequence 17, Appli
45	85	7.8	382	4	US-09-255-502-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-145-391-2
Sequence 2, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulist, Stefan M.
APPLICANT: Shibata, Hiroki
TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145,391
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 330
TYPE: PRT
ORGANISM: Homo sapiens
US-09-145-391-2

Query Match      50.7%; Score 551; DB 4; Length 330;
Best Local Similarity 97.1%; Pred. No. 2,6e-48;
Matches 102; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 MTNKKAVNPYTNQWKLNPVGVAVSPDFYAGTIVLLCQANQEGSSMYSGPSSLVYTSAMP 60
DB      196 MTNKKAVNPYTNQWKLNPVGVAVSPDFYAGTIVLLCQANQEGSSMYSPSSLVYTSAMP 255

QY      61 PPYPATTAAYRGHLEGRGRTVYNTFRRAAPPPIRAGGVY 105
DB      256 PPYPATTAAYRGHLEGRGRTVYNTFRRAAPPPIRAGGVY 300

RESULT 2
US-09-176-657-2
Sequence 2, Application US/09176657
Patent No. 6020164
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Tang, Y. Tom
APPLICANT: Corley, Neil C.
APPLICANT: Guejter, Karl J.
APPLICANT: Lu, Aina
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
FILE REFERENCE: PF-0611 US
CURRENT APPLICATION NUMBER: US/09/176,657
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
```

```

; SEQ ID NO 2
; LENGTH: 366
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1250374
US-09-176-657-2

```

Query Match	26.5%	Score 288;	DB 3;	Length 366;
Best Local Similarity	48.9%;	Pred. No. 1.9e-21;		
Matches 68;	Conservative 11;	Mismatches 54;	Indels 6;	Gaps 4

0y 1 MTNKAANPYNMGMLNUNVAVAVSPDEFAGTVLLCQAN--QEISSVYGGSSLD--VYTS 56

Db 190 MTNKKMTVPYANGMKLISVVGAVGVPBELIYAASSFOADVSLGNDAAVPLSGRGGINITYIPL 249

0y 57 AMPGEPY- AATAAANAAYGALHGRGRVTYNTFRAAAPPEIPAYGVGVYQEPYVGNKL 115

Db 250 IIPGPYPTAALTAAAFGAHLRGRGRITVGAVR- AVPTAIPAYPGVDMPIDMHSLL 308

0y 116 QGVAAVRYAQPATATAA 134

Db 309 QQPPLLPDLPDLPVTYMA 327

RESULT 3
US-08-195-152-2
; Sequence 2, Application US/08195152

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ADDRESSEE: Flehr, Hohnsch, Test Abritton & Hebert
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

```

1 APPLICATION NUMBER: US/08/195,152
2
3 FILING DATE: 14-FEB-1994
4
5 CLASSIFICATION: 435
6
7 ATTORNEY/AGENT INFORMATION:
8
9 NAME: Treacartin, Richard F.
10 REGISTRATION NUMBER: 31,801
11 REFERENCE/DOCKET NUMBER: A-59551/REF/RMS
12
13 TELECOMMUNICATION INFORMATION:
14
15 TELEPHONE: (415) 781-1989
16
17 TELEFAX: (415) 398-3249
18
19 TELEX: 910 277229
20
21 INFORMATION FOR SEQ ID NO: 2:
22
23 SEQUENCE CHARACTERISTICS:
24
25 LENGTH: 760 amino acids
26
27 TYPE: amino acid
28
29 TOPOLOGY: linear
30
31 MOLECULE TYPE: protein
32
33 US-08-195-152-2

```

```

Query Match      8.8%; Score 96; DB 1; Length 760;
Best Local Similarity 23.7%; Pred. No. 0.18;
Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6

QY 42 GSSMW-----SCPSSLVYTSAMPFPYPAATAAAVRGASHLRGRCTVYNTFRRAA 92

```

D _b	223	GSNLVYGCSSASNPIDGGVAIVNSSAV-----AAAAAAVVDGKH----	DYYYYNSMQYT	272
Q _y	93	PPPEPPIAYGVSVIYDEPPIYGNKLLOGGYAATRYAOPTPATAAA-----		134
D _b	273	PPPFYSGIYGTLYAATAATARAQAKMEPGAAAAAAAALITESTSYAASGNNNSQLYSSPYAGYNF		332
Q _y	135	-----YSDSYGRVYAADPETHHTLAPAP-----TYGVG-AMNFAFDLPDAKTRSHA		178
D _b	333	GQQDYGVGYINQIGNYISFPANYSRYAVSSPSBSMSHGHCFFHVAASSLSSESPDITHS		389

```

RESULT 4 4
US-09-343-011B-1
: Sequence 1, Application US/09343011B
: Patent No. 6300473
: GENERAL INFORMATION:
: APPLICANT: Stephanie Richard
: TITLE OF INVENTION: SLM-1 AND SLM-2; NOVEL
: FILE REFERENCE: A32561
: CURRENT APPLICATION NUMBER: US/09/343,011B
: PRIOR FILING DATE: 2001-05-11
: PRIOR APPLICATION NUMBER: CA 2265271
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: Fastrseq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 349
: TYPE: PRT
: ORGANISM: Mus musculus
US-09-343-011B-1

```

Query Match	8.7%;	Score 94;	DB 4;	Length 349;
Best Local Similarity	30.7%;	Pred. No. 0.1;		
Matches	54;	Conservative	11;	Mismatches
			51;	Indels
				60;
				Gaps
				12

OY	73	RGALHRRGRIRVYNTF	AAAPRPPIYAGGVVIOEPYNGKLQ	117
Db	186	RGRGRGRGIRITPTAASRCGCGAVPPPPRGRVLNPR	GTTVRGALPVPRIAGV	242
OY	118	-----GVA--YRYOOPRTAA-----	AYSDSYGRVYADPY--HHTLAPPT	157
Db	243	PTPRARGAIAIPGR-APPPAHAYEYEGYDDOYGEYEDDQTYEADNSVYPTQSVPE	307	
OY	158	---YGVG---ANNAFAPLIDAKTRSHADVDGLVLSLQ	---SYCGGNRRFPY	203
Db	302	YYDGHGHNEDAYSVAPEEWATRR	---SSLKAPPPRSAGGY-REHY	346

```

RESULT 5
US-09-252-292C-27
: Sequence 27, Application US/09252292C
: Patent No. 6245968
: GENERAL INFORMATION:
: APPLICANT: Boudec, Philippe
: APPLICANT: Rodgers, Matthew
: APPLICANT: Dumas, Florence
: TITLE OF INVENTION: Mutated hydroxyphenylpyruvate dioxygenase, DNA
: TITLE OF INVENTION: sequence and isolation of plants which contain such a
: FILE REFERENCE: 5500*31
: FILE REFERENCE: gene and which are tolerant to herbicides
: CURRENT APPLICATION NUMBER: US/09/252,292C
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 08/982,772
: PRIOR FILING DATE: 1997-12-02
: NUMBER OF SEQ. ID NOS: 31
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 27
: LENGTH: 444
: TYPE: DRT
: ORGANISM: Zea mays
US-09-252-292C-27

```


Query Match 8.6%; Score 93; DB 4; Length 444;
Best Local Similarity 22.8%; Pred. No. 0.18;
Matches 51; Conservative 22; Mismatches 69; Indels 82; Gaps 10;

QY 55 TSAMPGPYPATATAAARGAHLRG--RGRTVYTFPFAAAPPPIPAYGVY-----V 104
DB 112 TAAIPSFSAARFADHGLAVRAVALRAADAEAPFASVAGARPAFGPDLGRGRL 171
QY 105 YQEVNKGKLLGGGAAARVAOPTATAAAYSDSYGRV-----YAADPYHH-----T 151
DB 172 AEVELYGVULR--YVSY---PDGAGEPFLPGFEGVASPGAADYGLSRFDHIVGNVE 225
QY 152 LAPAPTY-----GVGAM-----NAFAPLTD-----AKTRSH 177
DB 226 LAPAAAFAGTGTGHEFAEFTEDVGTASGLNSVVLANNSENVLLPINEBVIHGRKRSQ 285
QY 178 -----ADVGLVLSLQASTYOGSYNRPAP 202
DB 286 IQTFLDHKGPGVQHMALASDDVLRITLREMOARSAMGGEFFMAP 329

RESULT 6
US-08-437-027-21
; Sequence 21, Application US/08437027
; Patent No. 5670317
; GENERAL INFORMATION:
; APPLICANT: Landanyi, Marc
; TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
; TITLE OF INVENTION: SMALL ROUND CELL TUMOR
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,027
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 46416/JPW/CCA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-437-027-21

Query Match 8.5%; Score 92.5; DB 1; Length 362;
Best Local Similarity 27.9%; Pred. No. 0.15;
Matches 48; Conservative 10; Mismatches 73; Indels 41; Gaps 9;

QY 29 YAGTVLLCOA-NOEGSSWYSGPSSLVYTSAMPGFYPATATAAARGAHLRGRTVYNT 87
DB 29 YAGTT---QAYGQGSYGYTGQPTDVSYTAQTTATYGTATATSY-----GQPTGYTT 79
QY 88 FRAAAPPPIPAYGVVQEPYVGNKLLGGYAAARYAOPTPATAAAYSDSYGRVYADP 147
DB 88 FRAAAPPPIPAYGVVQEPYVGNKLLGGYAAARYAOPTPATAAAYSDSYGRVYADP 147

DB 80 -----PTAPQ-----AYSQPVQGY-----GTGAYDTTATVTTQA-----SYAQS 116
QY 148 YHHTLAPAPTYGVGAMNAFAPLTDKTRSHADVGLVLSLQASTYOGSYNR 199
DB 117 AYGTQPAYPAYGQ-----QPAATAPTRPDGKNKPTETSPQSS--TGGVNO 160

RESULT 7
US-08-437-027-20
; Sequence 20, Application US/08437027
; Patent No. 5670317
; GENERAL INFORMATION:
; APPLICANT: Landanyi, Marc
; TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
; TITLE OF INVENTION: SMALL ROUND CELL TUMOR
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,027
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 46416/JPW/CCA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-437-027-20

Query Match 8.5%; Score 92.5; DB 1; Length 365;
Best Local Similarity 27.9%; Pred. No. 0.15;
Matches 48; Conservative 10; Mismatches 73; Indels 41; Gaps 9;

QY 29 YAGTVLLCOA-NOEGSSWYSGPSSLVYTSAMPGFYPATATAAARGAHLRGRTVYNT 87
DB 29 YAGTT---QAYGQGSYGYTGQPTDVSYTAQTTATYGTATATSY-----GQPTGYTT 79
QY 88 FRAAAPPPIPAYGVVQEPYVGNKLLGGYAAARYAOPTPATAAAYSDSYGRVYADP 147
DB 80 -----PTAPQ-----AYSQPVQGY-----GTGAYDTTATVTTQA-----SYAQS 116
QY 148 YHHTLAPAPTYGVGAMNAFAPLTDKTRSHADVGLVLSLQASTYOGSYNR 199
DB 117 AYGTQPAYPAYGQ-----QPAATAPTRPDGKNKPTETSPQSS--TGGVNO 160

RESULT 8
US-08-343-443B-2
; Sequence 2, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain

```

; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Meloc, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploougaestel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989,6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 656 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-343-443B-2

Query Match      8.5%; Score 92.5; DB 2; Length 656;
Best Local Similarity 27.9%; Pred. No. 0.33;
Matches 48; Conservative 10; Mismatches 73; Indels 41; Gaps 9;

```

```

QY 29 YAGTVLLCOA-NOEGSSMTSGPSLVYTSAMGFPYPATTAATAAAYGAHLRGGRGVYNT 87
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 29 YAGTT---QAYGQSGYGTGQPTDVSYTOAQTATYGTATVTSY-----GQPTGYTT 79
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 88 FRAAAPRPPIPAYGVVQEPYVGNKLLGGYAAVRYAOPTPATAAAGDSYRVYAADP 147
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 80 -----PTAPQ---AYSQPVQGY-----GTGAYDTTATVTTTQA-----SYAAS 116
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 148 YHHTLAPAPTYGVGAMNAFAPLTDKTRSHADVGLVLSLQASIVGGYNR 199
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 117 AVGTQPAYVAYGQ-----QPAATAPTRPQDGNKPTETSQPSS--TGGYNQ 160
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

RESULT 9
US-09-214-564A-4
; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.

```

```

; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
; TITLE OF INVENTION: Elongation By HIV-1 TAT
; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-214-564A-4

```

```

Query Match      8.5%; Score 92.5; DB 4; Length 656;
Best Local Similarity 27.9%; Pred. No. 0.33;
Matches 48; Conservative 10; Mismatches 73; Indels 41; Gaps 9;

QY 29 YAGTVLLCOA-NOEGSSMTSGPSLVYTSAMGFPYPATTAATAAAYGAHLRGGRGVYNT 87
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 29 YAGTT---QAYGQSGYGTGQPTDVSYTOAQTATYGTATVTSY-----GQPTGYTT 79
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 88 FRAAAPRPPIPAYGVVQEPYVGNKLLGGYAAVRYAOPTPATAAAGDSYRVYAADP 147
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 80 -----PTAPQ---AYSQPVQGY-----GTGAYDTTATVTTTQA-----SYAAS 116
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 148 YHHTLAPAPTYGVGAMNAFAPLTDKTRSHADVGLVLSLQASIVGGYNR 199
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 117 AVGTQPAYVAYGQ-----QPAATAPTRPQDGNKPTETSQPSS--TGGYNQ 160
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

RESULT 10
US-09-343-011B-2
; Sequence 2, Application US/09343011B
; Patent No. 6300473
; GENERAL INFORMATION:
; APPLICANT: Stephane Richard
; TITLE OF INVENTION: SLM-1 AND SLM-2; NOVEL
; TITLE OF INVENTION: SAM68-LIKE MAMMALIAN PROTEINS
; FILE REFERENCE: A32561
; CURRENT APPLICATION NUMBER: US/09/343,011B
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: CA 2265271
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-343-011B-2

```

```

Query Match      8.4%; Score 91; DB 4; Length 346;
Best Local Similarity 33.3%; Pred. No. 0.2;
Matches 32; Conservative 8; Mismatches 46; Indels 10; Gaps 4;

QY 60 GPPYPAATTAAYRGAGHLRGSG-----RTVYNT-FRAAAPRPPIPAYGVVQEPYVGN 112
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 221 GPTTRGVST--RGPVSKRGGLLTPRARGVPTGVRPPPTGTETGEYDYDGS-YGT 277
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 113 KLLGGYAAVRYAOPTPATAAAYSDSYGRVYAADPY 148
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 278 AYDEQSYDSYDYSYSTPAQASADYYDYGHGLSEDAY 313
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

RESULT 11
US-09-442-100-8

```

Sequence 8, Application US/09442100
Patent No. 6359193
GENERAL INFORMATION:
APPLICANT: Xu, Tian
APPLICANT: Tao, Wufan
APPLICANT: Wang, Weiyl
APPLICANT: Zhang, Sheng
APPLICANT: Yu, Wan
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
TITLE OF INVENTION: GENES AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,100
CLASSIFICATION:
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/411,111
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6523-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 980 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-442-100-8

Query Match 8.1%; Score 87.5; DB 4; Length 980;
Best Local Similarity 25.4%; Pred. No. 1.8;
Matches 50; Conservative 21; Mismatches 81; Indels 45; Gaps 9;

QY 26 PDFAAGTVLLCQANQEGSMYSGPSLVTYSAMP-----GFPYPATAAAYRGAHLRG 81
DB 108 PSEFGTEBALPSYHQLGGANGYEGPAL--EEMPRQYLDLFG--AGAGTHGAOHQHP 162
QY 82 RTVNTFRRAAPPPPIP-----AYGTVVQEPVYG-----NKLLQGVAAHYVAQ--- 126
DB 163 PKGST--AVPESAHFPGTHGRHLLSEQGYGVGRSSFGQNTPPRAYSSMKAQGGP 220
QY 127 ----PTPATAAAYSDSYGRVYAADPYHHTLAPAPTYG--VGAMNAFPLTDATRSHAD 179
DB 221 PASLTFFPAHAGLVYASHHKPAATPGAPHLVLTGTGPTFTGESSAQVLAPSR----- 274
QY 180 DVGVLSSLOASIQGG 196
DB 275 -----NSLNDLYELG 285

RESULT 12
US-08-552-142A-11
Sequence 11, Application US/08552142A
Patent No. 569595
GENERAL INFORMATION:
APPLICANT: Weintraub, Harold M.

APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes
TITLE OF INVENTION: and Proteins
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/552,142A
FILING DATE: 02-NOV-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,312
REFERENCE/DOCKET NUMBER: FPCR-1-8933
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-225-0709
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-552-142A-11

Query Match 8.0%; Score 87; DB 1; Length 379;
Best Local Similarity 20.9%; Pred. No. 0.58;
Matches 38; Conservative 23; Mismatches 73; Indels 48; Gaps 6;

QY 14 WKLPVVGAVYSPDFYAGTVLLC-----QANQEGSMYSGP 49
DB 171 WALSEILRSGRPLVYVOTLCKGLSQPTTNLVAGCLQNSRFLTEQGRDGAHRRHGS 230
QY 50 SSLVYTSAMPGFYPAA-----TAAAYRGAHLRGGR-TVYNT-FRAAADPPPIPA 99
DB 231 GG--PRAHMPYRYPGCRGRTVPGAAMAAGARLRTHTGYCAVETLYAAAGGASPD 287
QY 100 YGVVYQEPVYGNLLQGGTAAYRYAOPTPATAAYSDSYGRVYAADPYHHTLAPAPTYG 159
DB 288 YNSSEYEGPLSPPLCLNGNPSLKQDSSPDHEKSYHS-----MHVSGCGSRNHG 336
QY 160 VG 161
DB 337 HG 338

RESULT 13
US-08-946-914-14
Sequence 14, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Retner L.

APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-1994
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-14

Query Match 8.0%; Score 86.5; DB 3; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.4;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

QY 8 NPYTNGMKLNPNVGAIVSPDFYAGTVLLCOANOEGSSMYSGPSSLYTSMAMP-----GF 61
DB 16 NPNPGWMP-----GA-----WGNQPGAGYPGAS---YPGAYPGQAPPGGY 53
QY 62 P-----YPAATAAAYRGALHGRGRTVYNTFRAAPPPPIPAVGGVVQEPVYGNKL 114
DB 54 PGQAPPSAVPGPTGPSATPGP-----TAPGAYPGPTAP---GAPPGQP----- 93

QY 115 LOGGYAAVRYAQPPTATAAAYSDSYGRVYAADPYHHTLAPAPTYGVGAMNAFAPLT 170
DB 94 --GGPGAY-----PSAPGAYPSAPGAYPATGPF-----GAPLT-----GPLT 127

RESULT 14
US-09-656-450-14
Sequence 14, Application US/09656450
Patent No. 6468768
GENERAL INFORMATION:
APPLICANT: NI, Jian
APPLICANT: Gentz, Retner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
FILE REFERENCE: 1488.0560003
CURRENT APPLICATION NUMBER: US/09/656,450
CURRENT FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/263,689
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: US 08/946,914
PRIOR FILING DATE: 1997-10-09
PRIOR APPLICATION NUMBER: US 60/028,093
PRIOR FILING DATE: 1996-10-09

NUMBER OF SEQ ID NOS: 60
SOFTWARE: Patentin version 3.0
SEQ ID NO 14
LENGTH: 262
TYPE: PRT
ORGANISM: Rat
US-09-656-450-14

Query Match 8.0%; Score 86.5; DB 4; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.4;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

QY 8 NPYTNGMKLNPNVGAIVSPDFYAGTVLLCOANOEGSSMYSGPSSLYTSMAMP-----GF 61
DB 16 NPNPGWMP-----GA-----WGNQPGAGYPGAS---YPGAYPGQAPPGGY 53
QY 62 P-----YPAATAAAYRGALHGRGRTVYNTFRAAPPPPIPAVGGVVQEPVYGNKL 114
DB 54 PGQAPPSAVPGPTGPSATPGP-----TAPGAYPGPTAP---GAPPGQP----- 93

QY 115 LOGGYAAVRYAQPPTATAAAYSDSYGRVYAADPYHHTLAPAPTYGVGAMNAFAPLT 170
DB 94 --GGPGAY-----PSAPGAYPSAPGAYPATGPF-----GAPLT-----GPLT 127

RESULT 15
US-08-658-136-5
Sequence 5, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-658-136-5

Query Match 8.0%; Score 86.5; DB 3; Length 4302;
Best Local Similarity 24.3%; Pred. No. 17;
Matches 55; Conservative 22; Mismatches 80; Indels 69; Gaps 10;

QY 26 PDEYAGTVLLCOANQEGSSMYSGPSSLVYTSAMPGFYPYAA-----TAAAYRG 74
DB 185 PDNSSGTV-----AAVSFAAHEGLQPEACSAFCFSTGGLAALSEQ 228
QY 75 AHLRGRTVYNTFR-----AAAPPPPIPAVGIVYQEPVY-----GNKLQGG 118
DB 229 WCLGGAQPSASFACLSLCGPPPPPAFTCRGPTLLQHVFPASPGATLVGPHGPLASGQ 288
QY 119 YAAVRYAOPTPATAAA--YSDSYGRVYAADP-----YHHTLAPAPTYGVGAMNA 165
DB 289 LAAFHIAAPLPVATRWDFGDSAEVDAAAGPAASHRYVLPGRYHTAVLA----LGAGSA 344
QY 166 FAPLTDAKTRSHADVGIVL-----SLQASIV-OGGYNRFAPY 203
DB 345 LIG-TDVQVEAAPALBLVCPSSVQSDSLDLSTIQNRGSGGLEBAAY 389

Search completed: March 13, 2003, 00:55:08
Job time : 20 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:21:36 ; Search time 20 Seconds
(without alignments)
428.027 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTNKKAVNPYTNQKLNVPV.....VLSSLQASLYOGGYNRFAPY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	100.0	203	US-09-809-545A-2	Sequence 2, Appl1
2	551	50.7	330	US-09-794-591-2	Sequence 2, Appl1
3	96	8.8	760	US-08-754-311B-2	Sequence 14, Appl1
4	86.5	8.0	262	US-10-235-674-14	Sequence 14, Appl1
5	86.5	8.0	262	US-09-263-689-14	Sequence 14, Appl1
6	86	7.9	245	US-09-796-858-20	Sequence 20, Appl1
7	85	7.8	275	US-09-976-414-7	Sequence 7, Appl1
8	85	7.8	275	US-09-060-854B-3	Sequence 3, Appl1
9	85	7.8	275	US-10-075-895-2	Sequence 2, Appl1
10	85	7.8	275	US-10-075-895-2	Sequence 2, Appl1
11	85	7.7	382	US-10-050-624-31	Sequence 31, Appl1
12	83.5	7.7	485	US-09-925-300-1664	Sequence 1664, Ap
13	83	7.6	450	US-09-765-111A-14	Sequence 14, Appl1
14	82.5	7.6	400	US-09-073-009-126	Sequence 126, App
15	82.5	7.6	400	US-09-793-106-126	Sequence 126, App
16	82.5	7.6	499	US-09-863-877-2	Sequence 2, Appl1
17	82.5	7.6	499	US-09-732-350-2	Sequence 2, Appl1
18	81	7.5	551	US-09-920-300A-1789	Sequence 1789, Ap
19	81	7.5	551	US-10-033-528-1789	Sequence 1789, Ap

20	80.5	7.4	550	9	US-09-976-740-47	Sequence 47, Appl1
21	80.5	7.4	550	12	US-10-023-529-47	Sequence 47, Appl1
22	80.5	7.4	550	12	US-10-023-523-47	Sequence 47, Appl1
23	80	7.4	347	10	US-09-844-353A-100	Sequence 100, App
24	80	7.4	450	10	US-09-765-111A-25	Sequence 25, Appl1
25	80	7.4	527	9	US-09-712-163-156	Sequence 156, App
26	80	7.4	529	9	US-10-046-420-2	Sequence 2, Appl1
27	80	7.4	639	10	US-09-782-906-5	Sequence 5, Appl1
28	80	7.4	1062	10	US-09-815-242-5111	Sequence 5111, Ap
29	79.5	7.3	171	10	US-09-925-297-659	Sequence 659, App
30	79	7.3	381	10	US-09-920-118-16	Sequence 16, Appl1
31	79	7.3	723	12	US-10-044-205A-32	Sequence 32, Appl1
32	79	7.3	726	12	US-10-044-205A-31	Sequence 31, Appl1
33	79	7.3	3798	9	US-10-014-717-6	Sequence 6, Appl1
34	78.5	7.2	465	9	US-10-092-063-39	Sequence 39, Appl1
35	78	7.2	275	10	US-09-837-235-17	Sequence 17, Appl1
36	78	7.2	275	10	US-09-060-854B-4	Sequence 4, Appl1
37	78	7.2	302	9	US-10-029-180-18	Sequence 18, Appl1
38	78	7.2	357	10	US-09-837-235-15	Sequence 15, Appl1
39	77.5	7.1	240	10	US-09-819-917-8	Sequence 8, Appl1
40	77.5	7.1	382	10	US-09-809-033A-3	Sequence 3, Appl1
41	77.5	7.1	382	10	US-09-809-033A-4	Sequence 4, Appl1
42	77.5	7.1	399	9	US-10-029-180-64	Sequence 64, Appl1
43	77.5	7.1	1094	9	US-09-712-363-287	Sequence 287, App
44	77	7.1	265	10	US-09-819-254-2	Sequence 2, Appl1
45	77	7.1	311	9	US-10-025-380-1061	Sequence 1061, Ap

ALIGNMENTS

RESULT 1
US-09-809-545A-2
Sequence 2, Application US/09809545A
Patent No. US20020110804A1
GENERAL INFORMATION:
APPLICANT: Stanton, Lawrence W.
APPLICANT: White, R. Tyler
TITLE OF INVENTION: SECRETED FACTORS
FILE REFERENCE: SCTOS.017A
CURRENT APPLICATION NUMBER: US/09/809, 545A
CURRENT FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 203
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-809-545A-2

Query Match 100.0%; Score 1086; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.3e-94;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTNKKAVNPYTNQKLNVPVGVAVYSPDFYAGTVLLCQANQCGSSMYSGPSSLVYTSAMPG 60
DB 1 MTNKKAVNPYTNQKLNVPVGVAVYSPDFYAGTVLLCQANQCGSSMYSGPSSLVYTSAMPG 60

QY 61 PPYPATTAAYRGAHLFGRGRYTYNTFRRAAPPPIPAAYGVVYQEPVYGNKLLQGGYA 120
DB 61 PPYPATTAAYRGAHLFGRGRYTYNTFRRAAPPPIPAAYGVVYQEPVYGNKLLQGGYA 120

QY 121 AYRYQPTPATTAAYSDSGRYVYAADPYHHTLAPATYGVGAMNAFAPLTDKTRSHADD 180
DB 121 AYRYQPTPATTAAYSDSGRYVYAADPYHHTLAPATYGVGAMNAFAPLTDKTRSHADD 180

QY 181 VGLVLSLQASLYOGGYNRFAPY 203
DB 181 VGLVLSLQASLYOGGYNRFAPY 203

RESULT 2
US-09-794-591-2

; Sequence 2, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Shiba, Stefan M.
; APPLICANT: Shiba, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-591-2

Query Match 50.7%; Score 551; DB 10; Length 330;
Best Local Similarity 97.1%; Pred. No. 3 4e-44;
Matches 102; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MTNKAANDYNTGKMLNPNVGAIVSPDYAGTVLLCOANQEGSSMYSGPSLLVTSAMP 60
Db 196 MTNKTAVNTYNTGKMLNPNVGAIVSPDYAGTVLLCOANQEGSSMYSAPSLLVTSAMP 255
Qy 61 PPYPATAAARGAHLRGRGVNTFPRAAPPPPIPAVG 105
Db 256 PPYPATAAARGAHLRGRGVNTFPRAAPPPPIPAVG 300

RESULT 3

US-08-754-311B-2
; Sequence 2, Application US/08754311B
; Patent No. US2002004221A1
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albrighton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/754,311B
; FILING DATE: 21-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/195,152
; FILING DATE: 14-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-754-311B-2

Query Match 8.8%; Score 96; DB 8; Length 760;
Best Local Similarity 23.7%; Pred. No. 0.44;
Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6;

Qy 42 GSSMY-----SGPSLLVTSAMPPEPPYPAATAAARGAHLRGRGVNTFPRAA 92
Db 223 GSNLYGSSASNPDLGAVANSSAV-----AAAAAVYDCKH---DYIYNSMOQYT 272
Qy 93 PPPPIPAVGAVYQEPYVGNKLLQGYAARYAQPPTATAA----- 134
Db 273 PPPFVSGVTPTAATAARQAAMERGAATAAAAYLTPTSAASGNNSQULYSFVAGYNNF 332
Qy 135 -----YSDSYGRVYAADPYHHTLAPAP---TYGVG-AMNAFAPLTDATKTRSHA 178
Db 333 GQDDYGVYNEQYGNVYSPANYSPYAVSPSSSSASHGHGFHVAASNLSESPDTDHS 389

RESULT 4

US-10-235-674-14
; Sequence 14, Application US/10235674
; Publication No. US20030040081A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galactin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.056004
; CURRENT APPLICATION NUMBER: US/10/235,674
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 09/656,450
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent version 3.0
; SEQ ID NO 14
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Rat
US-10-235-674-14

Query Match 8.0%; Score 86.5; DB 9; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.9;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

Qy 8 NPYTNGKMLNPNVGAIVSPDYAGTVLLCOANQEGSSMYSGPSLLVTSAMP-----GF 61
Db 16 NNPNQGW-----GA-----WGNQPGAGVPGAS---YCGAYGQAAPG 53
Qy 62 P-----YPAATAAARGAHLRGRGVNTFPRAAPPPPIPAVGAVYQEPYVGNK 114
Db 54 PQQAPPSAIPGPTGPSAYVGP-----TABGAYFGPAP---GAFPGQP----- 93
Qy 115 LQGYAARYAQAQPTATAAAYSDSYGRVYAADPYHHTLAPATYGVGAMNAPAPLT 170
Db 94 --GGPGAY-----PSAPGAYPSAPGAYPATGPF-----GAPL-----G 127

RESULT 5

US-09-263-689-14
; Sequence 14, Application US/09263689
; Patent No. US20020150970A1
; GENERAL INFORMATION:

APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,689
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/946,914
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-263-689-14

Query Match 8.0%; Score 86.5; DB 10; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.9;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

QY 8 NPYNGMKLNIVGAVSPDFYAGVLLIQANOGSSMYSGPSSLYVTSAMP-----GF 61
DB 16 NPNQGNP-----GA-----WGNQPGAGTPGAAS--YGAIFGQAPPGGY 53
QY 62 P-----YPATTAAYRGALHGRGRTVNTFRAPAPPPPIYAGGVVYQEPYGNKL 114
DB 54 PGQAPSAVPGPTGSAVPGP-----TAPGAVPGTAP---GAPPGP----- 93
QY 115 LOGGYAAYRVQPPATTAAYSDSYGRVYAADPHYHTLAPPTYGVGMNFAPILT 170
DB 94 --GGPGAV-----PSAPGAYPSAPGAYPATGPF-----GAPLT-----GPILT 127

RESULT 6
US-09-796-858-20
Sequence 20, Application US/09796858
Patent No. US20020055139A1
GENERAL INFORMATION:
APPLICANT: Holtzman, Douglas
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING PROGNOSTIC, DIAGNOSTIC,
FILE REFERENCE: 7853-226-999
CURRENT APPLICATION NUMBER: US/09/796,858
PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 09/223,094
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/223,546
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/224,246

PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/312,359
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/336,536
PRIOR FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: 09/342,687
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 09/399,723
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 09/471,179
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 09/474,071
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/474,072
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/572,002
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: 09/597,993
PRIOR FILING DATE: 2000-06-12
PRIOR APPLICATION NUMBER: 09/599,596
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/606,565
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/365,164
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/630,334
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 09/665,666
PRIOR FILING DATE: 2000-09-20
NUMBER OF SEQ ID NOS: 50
SEQ ID NO 20
LENGTH: 245
TYPE: PRT
ORGANISM: Homo sapiens
US-09-796-858-20

Query Match 7.9%; Score 86; DB 10; Length 245;
Best Local Similarity 24.5%; Pred. No. 0.92;
Matches 52; Conservative 9; Mismatches 51; Indels 100; Gaps 11;

QY 26 PDFYAGT-----VLLQANQF-----GSSMYSG-----PSSLYVTSA----- 57
DB 55 PDFCCGSSQYCCSDVLKXIQWNEMCPEPSSRSFSAHPETPQLQSALKYQSSLDSDN 114
QY 58 MPGRPYATTAAYRGALHGRGRTVNTFRAP-----GLTFVVFPIATIIYVCTSCCCCLYKMCRRPVPVSN 161
DB 115 MPGR-----GATYAT-----GLTFVVFPIATIIYVCTSCCCCLYKMCRRPVPVSN 161
QY 92 -----APPPIYAGGVVYQEPYGNKLLOGGYAAYRVYQPTPA-TAAAYSDSY 139
DB 162 TTTVHTAVYQPPVAVPSVPGPIYQ-----GYHMPPOPGMFAAPYPTQY 207
QY 140 GRVYAADP-----YHHTLAPA--PTYGVGMN 164
DB 208 PPYLAQPTGPPAYHETLAGASQPPYNPAYMD 239

RESULT 7
US-09-976-414-7
Sequence 7, Application US/09976414
Patent No. US2002015575A1
GENERAL INFORMATION:
APPLICANT: No. US2002015575A1regard-Madsen, Mads
APPLICANT: Larsen, Line
APPLICANT: Hansen, Peter
TITLE OF INVENTION: Subtilase Variants
FILE REFERENCE: 10081.200-US
CURRENT APPLICATION NUMBER: US/09/976,414
CURRENT FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 275

TYPE: PRT
ORGANISM: Bacillus
US-09-976-414-7

Query Match 7.8%; Score 85; DB 9; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKMLNPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSAAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALRGGRGRTVNTFRAAAPPPPIPAYGGVYQEPYGNKLLQGGYA 120
DB 169 GKYPVIAVGAVDSSNQRA-----SFSSVGPBLDVMA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPATAAAYSDSYGRVYAADPYHHTLAAPT-----YGVGMNFA 167
DB 219 GTSMASPHVGAALILSKHPMTNTQVRSSLENTTTKLGDSPFYGKGLINVOA 272

RESULT 8
US-09-060-854B-3
Sequence 3, Application US/09060854B
Patent No. US2002081703A1
GENERAL INFORMATION:
APPLICANT: Eceli, David Aaron
TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002008
FILE REFERENCE: GC532
CURRENT APPLICATION NUMBER: US/09/060,854B
CURRENT FILING DATE: 1998-04-15
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 275
TYPE: PRT
ORGANISM: B. amyloliquefaciens
US-09-060-854B-3

Query Match 7.8%; Score 85; DB 10; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKMLNPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSAAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALRGGRGRTVNTFRAAAPPPPIPAYGGVYQEPYGNKLLQGGYA 120
DB 169 GKYPVIAVGAVDSSNQRA-----SFSSVGPBLDVMA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPATAAAYSDSYGRVYAADPYHHTLAAPT-----YGVGMNFA 167
DB 219 GTSMASPHVGAALILSKHPMTNTQVRSSLENTTTKLGDSPFYGKGLINVOA 272

RESULT 9
US-10-075-907-2
Sequence 2, Application US/10075907
Patent No. US20020127695A1
GENERAL INFORMATION:
APPLICANT: Genencor International, Inc.
APPLICANT: Davis, Benjamin G.
APPLICANT: Jones, Bryan J.
APPLICANT: Bott, Richard R.
TITLE OF INVENTION: CHEMICALLY MODIFIED ENZYMES WITH
FILE REFERENCE: 23623-7039
CURRENT APPLICATION NUMBER: US/10/075,907
CURRENT FILING DATE: 2002-02-12
PRIOR APPLICATION NUMBER: US 60/113,130

PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 275
TYPE: PRT
ORGANISM: Bacillus amyloliquefaciens
FEATURE:
OTHER INFORMATION: Subtilisin BPN'
US-10-075-907-2

Query Match 7.8%; Score 85; DB 12; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKMLNPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSAAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALRGGRGRTVNTFRAAAPPPPIPAYGGVYQEPYGNKLLQGGYA 120
DB 169 GKYPVIAVGAVDSSNQRA-----SFSSVGPBLDVMA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPATAAAYSDSYGRVYAADPYHHTLAAPT-----YGVGMNFA 167
DB 219 GTSMASPHVGAALILSKHPMTNTQVRSSLENTTTKLGDSPFYGKGLINVOA 272

RESULT 10
US-10-075-895-2
Sequence 2, Application US/10075895
Patent No. US2002013717A1
GENERAL INFORMATION:
APPLICANT: Genencor International, Inc.
TITLE OF INVENTION: MODIFIED ENZYMES AND THEIR USE FOR
TITLE OF INVENTION: PEPTIDE SYNTHESIS
FILE REFERENCE: 23623-7013
CURRENT APPLICATION NUMBER: US/10/075,895
CURRENT FILING DATE: 2002-02-13
PRIOR APPLICATION NUMBER: US 60/072,265
PRIOR FILING DATE: 1998-01-23
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 275
TYPE: PRT
ORGANISM: Bacillus amyloliquefaciens
FEATURE:
OTHER INFORMATION: Subtilisin BPN'
US-10-075-895-2

Query Match 7.8%; Score 85; DB 12; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKMLNPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSAAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALRGGRGRTVNTFRAAAPPPPIPAYGGVYQEPYGNKLLQGGYA 120
DB 169 GKYPVIAVGAVDSSNQRA-----SFSSVGPBLDVMA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPATAAAYSDSYGRVYAADPYHHTLAAPT-----YGVGMNFA 167
DB 219 GTSMASPHVGAALILSKHPMTNTQVRSSLENTTTKLGDSPFYGKGLINVOA 272

RESULT 11
US-10-090-624-31
Sequence 31, Application US/10090624
Patent No. US20020132335A1

GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, MIO
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 31
LENGTH: 382
TYPE: PRT
ORGANISM: Bacillus subtilis
US-10-090-624-31

Query Match 7.8%; Score 85; DB 12; Length 382;
Best Local Similarity 26.4%; Pred. No. 2;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYTNQMKLNPNVGVAVSPDFYAGTVLLCOANQSGSSMYSGPSLSLYTSAMPGF 62
DB 225 NMDVYNNSLGSPSSALKAADVAKVAGVYVVAAGNGBE-SSGSTY-GRF 275
QY 63 --YPAATAAAYRGAILRGRGRVYNTFRAPPPPIPAYGVVQEPYGNKLLGGYA 120
DB 276 GKYSVIAVAGVDSNQA-----SFSVGPFLDVMA-PGVSIQSTLPGNK--YGAYN 325
QY 121 AYRAQPPATAAAYSDSYGRVYADPHHTLAPPT-----YGVANMAFA 167
DB 326 GTSMASPHVAGAAALLSKHPMTNTQVRSSELENTTKLGDSPYKGLINVQA 379

RESULT 12
US-09-925-300-1664
Sequence 1664, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1664
LENGTH: 485
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-300-1664

Query Match 7.7%; Score 83.5; DB 10; Length 485;
Best Local Similarity 24.8%; Pred. No. 3.7;
Matches 39; Conservative 15; Mismatches 70; Indels 33; Gaps 8;

QY 9 PYTGMKLNPNVGVAVSPDFYAGTVLLCOANQSGSSMYSGPSLSLYTSAMPGFYPAATA 68
DB 52 PYPGGRF--PWGGAGYPP-----VSSGYPGAGGYPAGGCGYAPAGGYPGAG--G 99
QY 69 AAAYRGAILRGRGRVYNTFRAPPPPIPAYGVVQEPYGNKLLGGYAAYRY--A 125
DB 100 APYVGVPPGCGFVPPGAGFSYFPQPPSQSYGGGPAQVPLPG--PPGQMSQYFPG 157

QY 126 QPT-----PATMAAYSDSYGRVYADPHHTLAPPTY 158
DB 158 QPTYSQPATVTVTQGT-----TIRPANF 182

RESULT 13
US-09-765-111A-14
Sequence 14, Application US/09765111A
Patent No. US20020106796A1
GENERAL INFORMATION:
APPLICANT: Fletcher, Jonathan A.
APPLICANT: Kroll, Todd G.
TITLE OF INVENTION: PAX8-PARAGAMMA NUCLEIC ACID MOLECULES
FILE REFERENCE: B0801/7196/ERP/MAT
CURRENT APPLICATION NUMBER: US/09/765,111A
CURRENT FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: US 60/177,109
PRIOR FILING DATE: 2000-01-20
PRIOR APPLICATION NUMBER: US 60/225,079
PRIOR FILING DATE: 2000-08-14
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 450
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-765-111A-14

Query Match 7.6%; Score 83; DB 10; Length 450;
Best Local Similarity 26.5%; Pred. No. 3.8;
Matches 49; Conservative 12; Mismatches 68; Indels 56; Gaps 8;

QY 18 PVGAVVSPDFYAGTVLLCOANQSGSSMYSGPSLYVT-----SAMP--GFRYPA 65
DB 296 PVVADPHSP-----FAIKETPEVSSSTPSSLSAFLDLQVSGVPPFPAFPA 349
QY 66 ATAAAYRGAILRGR--GRVYNTFRAPPPPIPAYGVVQEPYGNKLLGGYAAY 122
DB 350 SVYQFTGQALLSRENVGPTL-----PGYPRHPTSG-----QGSYAS- 388
QY 123 RYAPPTATAAAYSDSYGRVYADPY-----HHTLAPPTYGVANMAFAPLTD 172
DB 389 ---SALAGWAGSEYSGNAVYHTPYSSSEAWRFNSSLSPYYSSTSRPAPPTA 444
QY 173 KTRSH 177
DB 445 TAFDH 449

RESULT 14
US-09-073-009-126
Sequence 126, Application US/09073009
Patent No. US20010012888A1
GENERAL INFORMATION:
APPLICANT: Alderson, Mark
APPLICANT: Dillon, Davin C.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

TUBERCULOSIS AND MEJ

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-632-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 126:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-073-009-126

Query Match 7.6%; Score 82.5; DB 10; Length 400;
Best Local Similarity 26.1%; Pred No. 3.6;
Matches 48; Conservative 18; Mismatches 81; Indels 37; Gaps 8;

QY 34 LLCQANQEGSSMYSGPSSLVYTSAMPFPYPATATAAAYRGHILRGRTVYNTFRPA-- 91
DB 48 VITQSSSG--WLGPAASAMAEVA--PYVAMMSAAAQAEQATQRAAAAAFEAAFA 102
QY 92 --APPPIPA-----YG---GVYQEPVYGNKLLGGYAAVRYAQTPTPATA 132
DB 103 ATVPPLIAANRASIMQLISTNVFGQNTSALAAAEQYGEWMAQDSAAAYAVAGSSASAS 162
QY 133 AAYSDSYGRVYAADPYHHTLAPARTYGVGAMNAFAPLTDKTRSHADVGL--VLSLQA 190
DB 163 AVTPSTPPQIANPTAQGTQAAVAATAAG--TAOSTLTEMIT-----GLPNALQSILTS 213
QY 191 SIYQ 194
DB 214 XLQ 217

RESULT 15
US-09-793-306-126
Sequence 126, Application US/09793306
Patent No. US20020098200A1
GENERAL INFORMATION:
APPLICANT: Campos-Neto, Antonio
APPLICANT: Skeiky, Yaelir
APPLICANT: Ovendale, Pamela
APPLICANT: Jen, Shylan
APPLICANT: Lodes, Michael
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
FILE REFERENCE: 014058-008740US
CURRENT APPLICATION NUMBER: US/09/793,306
CURRENT FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 60/185,037
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/223,828
PRIOR FILING DATE: 2000-08-08
NUMBER OF SEQ ID NOS: 164
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 126
LENGTH: 400
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: Tb470 (Mtb-40)
NAME/KEY: MOD_RES
LOCATION: (1)..(400)
OTHER INFORMATION: Xaa = any amino acid

US-09-793-306-126
Query Match 7.6%; Score 82.5; DB 10; Length 400;
Best Local Similarity 26.1%; Pred. No. 3.6;
Matches 48; Conservative 18; Mismatches 81; Indels 37; Gaps 8;

QY 34 LLCQANQEGSSMYSGPSSLVYTSAMPFPYPATATAAAYRGHILRGRTVYNTFRPA-- 91
DB 48 VITQSSSG--WLGPAASAMAEVA--PYVAMMSAAAQAEQATQRAAAAAFEAAFA 102
QY 92 --APPPIPA-----YG---GVYQEPVYGNKLLGGYAAVRYAQTPTPATA 132
DB 103 ATVPPLIAANRASIMQLISTNVFGQNTSALAAAEQYGEWMAQDSAAAYAVAGSSASAS 162
QY 133 AAYSDSYGRVYAADPYHHTLAPARTYGVGAMNAFAPLTDKTRSHADVGL--VLSLQA 190
DB 163 AVTPSTPPQIANPTAQGTQAAVAATAAG--TAOSTLTEMIT-----GLPNALQSILTS 213
QY 191 SIYQ 194
DB 214 XLQ 217

Search completed: March 13, 2003, 00:53:19
Job time : 21 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus v2n model

Run on: March 16, 2003, 03:31:48 ; Search time 1706 Seconds
(without alignments)
2524.745 Million cell updates/sec

Title: US-09-809-545a-2_COPY_1_148
Perfect score: 148
Sequence: 1 MTKKAVNPYTNCKMLNPV.....PATTAAYSDSYGRVYADPY 148

Scoring table: BLIGO
Xgapop 60.0 , Xgapext 60.0
Xgapop 60.0 , Xgapext 60.0
Xgapop 6.0 , Xgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgnr_1/USPNO.spool/US09809555/runat_07032003_151954_12159/app_query.fasta.1.327
-DB=genmbd1 -QPM=fastlap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRL=US09809555 @CGN 1 1 3745 @runat_07032003_151954_12159 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAT -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOS -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=6 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Genemdb1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rdd:*
36: em_hcg_man:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	1340	6 AX268800	AX268800 Sequence
2	100	67.6	1363	10 AF107204	AF107204 Mus muscu
3	100	67.6	1547	9 AF094849	AF094849 Homo sapi
4	100	67.6	2000	9 AF229057	AF229057 Homo sapi
5	100	67.6	2002	10 AB041596	AB041596 Mus muscu
6	89	60.1	3348	9 AF109106	AF109106 Homo sapi
7	60	40.5	1586	10 AF191501	AF191501 Mus muscu
8	58	39.2	1475	9 AB060859	AB060859 Macaca fa
9	58	39.2	1513	9 AK001027	AK001027 Homo sapi
10	58	39.2	2279	9 AF107203	AF107203 Homo sapi
11	57	38.5	2372	6 AR134676	AR134676 Sequence
12	44	29.7	251	9 HSF0X09	AF109115 Homo sapi
13	44	29.7	164538	9 AC005774	AC005774 Homo sapi
14	34	23.0	596	9 HSA33368	AJ333688 Homo sapi
15	34	23.0	617	9 HSA335811	AJ335811 Homo sapi
16	34	23.0	653	9 HSA335616	AJ335616 Homo sapi
17	34	23.0	675	9 HSA334013	AJ334013 Homo sapi
18	21	14.2	151	9 HSF0X11	AF109117 Homo sapi
19	21	14.2	115860	2 AC120661	AC120661 Rattus no
20	20	13.5	1623	9 HSM803066	AF109114 Homo sapi
21	17	11.5	374	9 HSF0X08	AF109114 Homo sapi
22	16	10.8	959	10 AF229056	AF229056 Mus muscu
23	14	9.5	249	9 HSF0X10	AF109115 Homo sapi
24	14	9.5	347	9 HSF0X07	AF109113 Homo sapi
25	14	9.5	112976	9 AC027683	AC027683 Homo sapi
26	14	9.5	115860	2 AC120661	AC120661 Rattus no
27	13	8.8	1134	10 AF387322	AF387322 Mus muscu
28	13	8.8	1393	9 AY072786	AY072786 Homo sapi
29	13	8.8	1538	9 BC013115	BC013115 Homo sapi
30	13	8.8	1553	9 BC025281	BC025281 Homo sapi
31	13	8.8	1558	9 AK055213	AK055213 Homo sapi
32	13	8.8	1695	10 AF229055	AF229055 Mus muscu
33	13	8.8	1721	10 BC027263	BC027263 Mus muscu
34	13	8.8	1876	9 HSP18A20	AL009266 H. sapien
35	13	8.8	1925	9 AF229058	AF229058 Homo sapi
36	13	8.8	108315	9 HSA4182	AL049748 Human DNA
37	13	8.8	156549	2 AC079072	AC079072 Homo sapi
38	13	8.8	156820	2 AC123391	AC123391 Rattus no
39	13	8.8	159565	10 AL591075	AL591075 Mouse DNA
40	13	8.8	203649	10 AL603983	AL603983 Mouse DNA
41	13	8.8	208632	2 AC073624	AC073624 Homo sapi
42	13	8.8	209461	9 AC079359	AC079359 Homo sapi
43	12	8.1	428	6 AX336745	AX336745 Sequence
44	12	8.1	444	9 HSA33571	AJ33571 Homo sapi
45	12	8.1	669	9 HSA338965	AJ338965 Homo sapi

RESULT 1

ALIGNMENTS

AX268800
LOCUS AX268800 1340 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1
AUTHORS Stanton, L.W. and White, R.T.
TITLE Secreted factors
JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scienc Inc. (US)
FEATURES
source location/Qualifiers
1..1340
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN
Alignment Scores:
Pred. No.: 8,28e-119 Length: 1340
Score: 148.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-09-809-545a-2_copy_1_148 (1-148) x AX268800 (1-1340)
QY 1 MetThrAsnLysLeuValAlaValaenProTyrThrAsnGlyTyrPheLeuAenProValaVal 20
DB 535 ATGACTAAATTAAGAGCCCGGAAACCCCTACACCAATGCGTGAATTAATTCAGATTG 54
QY 21 GAlaValaValTyrSerProAapPheTyrAlaGlyTyrValLeuLeuCyGlnAlaAenGln 40
DB 595 GCGCGCGTCTACAGCCCCGACTCTGTATGACGACGCGTGTGTGTGCGCAGCCACAG 654
QY 41 GAlaGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
DB 655 GAGGAGTCTTCCATGATGACAGTGGCCCCAGTCTGTATATATCTTGCATATGCTGGC 714
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyValAlaHisLeuArgGlyArg 80
DB 715 TTTCCATATCCGCGCCCGCACTGCTGACGCTGACATACGAGGGCTCACTTCAGGCCGT 774
QY 81 GAlaArgThrValTyrAanThrPheArgAlaAlaAlaProProProProLeuProAlaTyr 100
DB 775 GGTGCGACCGTGTACAAACCTTCAAGAGTGGCGGCCGCCCAACCCCAATCCCGGCTAT 834
QY 101 GAlaGlyValaValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlnGlyTyrAla 120
DB 835 GCGCGAGTGTATATAGAGCGCATGATGCGCAATTAATTTGCTAAGGGTGTTCAGCT 894
QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAapSerTyrGly 140
DB 895 GCATACCGCTACGCCCGACGCCCGCCCTGCTGCTGCTTACATGATGATGATGCGA 954
QY 141 ArgValTyrAlaAlaAapProTyr 148
DB 955 CGAGTTTATGCTGCCGACCCCTAC 978
RESULT 2
AF107204 AF107204 1363 bp mRNA linear ROD 23-JUN-2000
DEFINITION Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS
SOURCE Mus musculus.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1363)
AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulet, S.-M.
TITLE A novel protein, A2BP, with RNA binding motif, binds to C-terminal
ataxin-2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1363)
AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulet, S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA
FEATURES
source location/Qualifiers
1..1363
/organism="Mus musculus"
/db_xref="taxon:10090"
1..1363
/gene="A2bp"
67..1203
/gene="A2bp"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AA178292.1"
/db_xref="GI:8671588"
/translation="MAQPYASQAPAPONGIAPAYTAPHPAPAEYTGQTTPDHTLN
LYPTQHSBOSADTSAGTASGTATOTDAAPTGQPTQSPENTSESOPKRLVSN
IPFRFRPDPLROMFOGFGKILDEIIFNERSGSGRPFPENASADBRERKLTCT
VEGSKLEVNNAATVMTNKTVPNTPTNGKLNIPVGAIVYSPDPAGVYLCOANOEG
SSWYSGPSLIVYTSAMPGEYPAPATAAAYGAHLRGKRTVYNTFRRAAAPPPIPAY
GVVYODGFYADIVGGYAAVRYAQPTPATAAASDSYGRVYADPFRHHTLAAPTGY
VGANNAFAPFLDADKTRSHADVDVGLSLQASIVRGYNRAPAY"
358..375
/gene="A2bp"
/note="Region: RNA binding motif RNP-2"
misc_feature
475..495
/gene="A2bp"
/note="Region: RNA binding motif RNP-1"
misc_feature
/note="Region: RNA binding motif RNP-1"
BASE COUNT 362 a 396 c 324 g 281 t
ORIGIN
Alignment Scores:
Pred. No.: 2,64e-77 Length: 1363
Score: 100.00 Matches: 100
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.57% Indels: 0
DB: Gaps: 0
US-09-809-545a-2_copy_1_148 (1-148) x AF107204 (1-1363)
QY 7 ValaenProTyrThrAsnGlyTyrPheLeuAenProValaValAlaValaValTyrSerPro 26
DB 613 GTCAACCCCTTACCAACATGCGTGAATTAATTCAGTTGTGGCGGCTGTACAGCCCC 672
QY 27 AapPheTyrAlaGlyTyrValLeuLeuCyGlnAlaAenGlnGlySerSerMetTyr 46
DB 673 GACTTCTATGACGCGCGTGTGTTGCGCAGGCCACCAAGAAGGATCTTCATGTAC 732
QY 47 SerGlyProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAla 66
DB 733 AGTGGCCCGCATCTTATATATCTTGTGCAATGCCCTGCTCCCATATCCGGCGCC 792
QY 67 ThrAlaAlaAlaAlaTyrArgGlyValaHisLeuArgGlyArgGlyArgThrValTyrAan 86
DB 793 ACTGTGCGAGCTGCATACCGAAGGGCTCACCTTCAAGGCGGTGTCACCGTATCAAC 852
QY 87 ThrPheArgAlaAlaAlaProProProProLeuProAlaTyrGlyGlyValaValTyrGln 106
DB 853 ACCTTCAGGCTGACAGCGCCCGCCCGCCCAATCCGCGCTATGCGGCTGTGTTTACAG 912

```
RESULT 3
AF094849      1547 bp  mRNA      linear  PRI 01-MAR-2002
LOCUS
DEFINITION   Homo sapiens hexaribonucleotide binding protein 1 isoform alpha
              (HRNBp1) mRNA, complete cds.
ACCESSION    AF094849
VERSION      AF094849.1  GI:19032365
KEYWORDS
SOURCE       Homo sapiens.
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 1547)
AUTHORS      Chen, W., Chu, Z.-L., Blough, R. I., Liu, L., Hoppes, B. and
              Winkelman, J. C.
TITLE        Molecular cloning and chromosomal localization of a human brain,
              heart and skeletal muscle specific RNA binding protein gene
              homologous to fox-1 in Caenorhabditis elegans
              Unpublished
              2 (bases 1 to 1547)
JOURNAL
REFERENCE    Submitted (24-SEP-1998) Internal Medicine/Hematology-Oncology,
AUTHORS      University of Cincinnati College of Medicine, 231 Bethesda Ave.,
              Cincinnati, OH 45267-0508, USA
              Location/Qualifiers
FEATURES
source       1. 1547
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="16"
              /map="16p13.3"
              1. 1547
              /gene="HRNBp1"
              254. 1432
              /gene="HRNBp1"
              /note="similar to Caenorhabditis elegans fox-1"
              /codon_start=1
              /product="hexaribonucleotide binding protein 1 isoform
              alpha"
              /protein_id="AA183405.1"
              /db_xref="GI:19032366"
              /translation="MLASQGVLLHPYGVPMIVPAPYLPGLIQNOEAAAAPDTMAOP
              VASQAPRPNQGI PAEYTAHPHAPRYGTQTVPEHTLNLVPAQTHSESPADTSA
              QTVSGTATQDDAAPTQDQPTQPSSENTENKSKPKLHVSNI PRFRDPLRQMGQF
              KTLIDVEIITNERSKSGFGVTPENSADARARKLHGTVEGKIEVNNATAVMTN
              KKTVPNTNGKLNIPVGVAVSPFVAGTVLLCOANQGSMSVAPSLSVTSAMPF
              PYPAATAAARGLRGRGTIVNTFRAAAPPIPIAYGGVVQEPVYGNKLQGGY
              AAYRYQPTPATAAVSDSYGRVYAADPYHHLAPAPTYGVGMASIVRGYNRFAY
              "
BASE COUNT   380 a      448 c      374 g      345 t
ORIGIN
Alignment Scores:
Pred. No.:   2,91e-77      Length:      1547
Score:        100.00      Matches:      100
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:  67.57%      Indels:      0
              Gaps:      0
US-09-809-545A-2_COPY_1_148 (1-148) x AF094849 (1-1547)
QY 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
DB 1043 CCCAGTTCACTTGATATATACCTTCGCAATGCCAGGCTTCCCTATCCAGCACCCGCC 1102
QY 69 AAlaAlaIaIaTyrArgGlyAlaHisIeuaRgGlyArgGlyArgThValTyrAsnThrPhe 88
DB 1103 GCGGCGGCTACCGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1162
QY 89 ATGAlaAlaIaIaProProProProIleProAlaTyrGlyGlyValValTyrGlnGluPro 108
```

```
DB 1163 AGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1222
QY 109 ValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThr 128
DB 1223 GTGTATGGCAATTAATATGCTGCAGAGGTGGTATGCTGATACCGCTACGCCAGCCCTAC 1282
QY 129 ProAlaThrAlaAlaAlaIaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspProTyr 148
DB 1283 CCTGCCACTCGCGCTGCTACGATGACGATGACGAGAGAGATTATGCTGCGACCCCTAC 1342
RESULT 4
AF229057      2000 bp  mRNA      linear  PRI 01-FEB-2002
LOCUS
DEFINITION   Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
              (HRNBp1) mRNA, complete cds.
ACCESSION    AF229057
VERSION      AF229057.1  GI:18461366
KEYWORDS
SOURCE       Homo sapiens.
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 2000)
AUTHORS      Chen, W. and Winkelman, J. C.
TITLE        Direct Submission
JOURNAL
REFERENCE    Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
AUTHORS      University of Cincinnati College of Medicine, The Vontz Center for
              Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
              Location/Qualifiers
FEATURES
source       1. 2000
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="16"
              /map="16p13.3"
              1. 2000
              /gene="HRNBp1"
              256. 1512
              /gene="HRNBp1"
              /note="RNA-binding protein"
              /codon_start=1
              /product="hexaribonucleotide binding protein 1 isoform
              gamma"
              /protein_id="AA171904.1"
              /db_xref="GI:18461367"
              /translation="MLASQGVLLHPYGVPMIVPAPYLPGLIQNOEAAAAPDTMAOP
              VASQAPRPNQGI PAEYTAHPHAPRYGTQTVPEHTLNLVPAQTHSESPADTSA
              QTVSGTATQDDAAPTQDQPTQPSSENTENKSKPKLHVSNI PRFRDPLRQMGQF
              KTLIDVEIITNERSKSGFGVTPENSADARARKLHGTVEGKIEVNNATAVMTN
              KKTVPNTNGKLNIPVGVAVSPFVAGTVLLCOANQGSMSVAPSLSVTSAMPF
              PYPAATAAARGLRGRGTIVNTFRAAAPPIPIAYGGVVQEPVYGNKLQGGY
              AAYRYQPTPATAAVSDSYGRVYAADPYHHLAPAPTYGVGMASIVRGYNRFAY
              ADDVGLVSSLSQASIVRGYNRFAY"
BASE COUNT   502 a      505 c      487 g      506 t
ORIGIN
Alignment Scores:
Pred. No.:   3,55e-77      Length:      2000
Score:        100.00      Matches:      100
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:  67.57%      Indels:      0
              Gaps:      0
US-09-809-545A-2_COPY_1_148 (1-148) x AF229057 (1-2000)
QY 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
DB 1045 CCCAGTTCACTTGATATATACCTTCGCAATGCCAGGCTTCCCTATCCAGCACCCGCC 1104
QY 69 AAlaAlaIaIaTyrArgGlyAlaHisIeuaRgGlyArgGlyArgThValTyrAsnThrPhe 88
DB 1105 GCGGCGGCTACCGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1164
```

Oy	69	AGGAAAGAATGACCTGGTGTTCGCGCAGGTTATGCCTACGCCAGCCACTACC	1284
Oy	109	GATTATGCAATTAATGCTGCGAGGTTATGCTGCGATACGCCAGCCAGCTACC	1284
Oy	129	ProAlaThAlAAlAAlAtYrSeRaSpSrTYrGYaRgValTYrAlAAlAAsPPrOTr	148
Db	1285	CCTGCACACTCCGCGCTCACAGTGACAGTTATGCAGCATTCGCGACCCCTAC	1344
AB041596			
LOCUS	AB041596	2002 bp	mRNA linear ROD 30-JUN-2000
DEFINITION	Mus musculus brain cDNA, clone MNcb-3035, similar to Homo sapiens		
VERSION	AB041596		
KEYWORDS	AB041596.1 GI:7670455		
SOURCE	file (full insert sequence).		
ORGANISM	Mus musculus (strain:C57BL/6J) adult female cDNA to mRNA,		
REFERENCE	clone lib:Sugano mouse brain mncb clone:MNCB-3035.		
AUTHORS	Mus muscularis		
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	Oseada,N., Kusuda,J., Tanuma,R., Ito,A., Hirata,M., Sugano,S. and		
AUTHORS	Hashimoto,K.		
TITLE	Isolation of full-length cDNA clones from mouse brain cDNA library		
JOURNAL	made by oligo-capping method		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 2002)		
TITLE	Hashimoto,K., Oseada,N., Kusuda,J. and Sugano,S.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of		
AUTHORS	Infectious Diseases, Division of Genetic Resources; 23-1, Toyama		
TITLE	1-chome, Shinkuku-Ku, Tokyo 162-8640, Japan		
JOURNAL	(E-mail:khashim@nih.go.jp), URL:http://www.nih.go.jp/yoken/genebank/,		
REFERENCE	Tel.:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)		
AUTHORS	URL: http://www.nih.go.jp/yoken/genebank/		
TITLE	lib Name: Sugano mouse brain mncb		
JOURNAL	Lab host: TOB10		
REFERENCE	Vector: pME18S-FU3		
AUTHORS	1st strand cDNA was primed with an oligo(dT) primer		
TITLE	[ATCGGCGCTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized		
JOURNAL	using specific 5' and 3' primers and amplified by PCR. The PCR		
REFERENCE	product was digested with SfiI and size selection was performed to		
AUTHORS	exclude fragments <1.5kb. The SfiI-digested PCR product was cloned		
TITLE	into distinct DraIII sites of pME18S-FU3. XhoI sites just outside		
JOURNAL	the DraIII sites can be used to isolate the cDNA insert. Library		
REFERENCE	was constructed by Sugano et al. (University of Tokyo, Institute of		
AUTHORS	Medical Science). Custom primer used for sequencing (5' end primer		
TITLE	[CTTTCGCTCTMAAAGCTGGG]; 3' end primer		
JOURNAL	[CGACCTGCAGCTCGACGACA]);		
REFERENCE	A part of this sequence is reported in AU067167.		
AUTHORS	Location/Qualifiers		
TITLE	1..2002		
JOURNAL	/organism="Mus musculus"		
REFERENCE	/strain="C57BL/6J"		
AUTHORS	/db_xref="taxon:10090"		
TITLE	/clone="MNcb-3035"		
JOURNAL	/sex="female"		
REFERENCE	/clone_lib="Sugano mouse brain mncb"		
AUTHORS	/dev_stage="adult"		
TITLE	637..1827		
JOURNAL	/note="unnamed protein product"		
REFERENCE	/codon_start=1		
AUTHORS	/protein_id="BA095079.1"		
TITLE	/db_xref="gi:7670456"		
JOURNAL	/translation="MNCERDRLRGNOFAAADPTMAOPYASQAFAPPONGIPAYPAR		
REFERENCE	HPHAPETVTCGTTVGHTLNLTPPTQTHSEBSADTSATVSCTATOTDDAAPDTGQQP		
AUTHORS	TQPSNDSQPRLHVSNIVPRFRDLQOMGFQFKILIDVEITFNERKSKSGFVW		
TITLE	TFENSADDDRAREKLGHTGVNIGRKLEVNNAARWTNNKTVPNTNGMKLNPPVGAAYV		

BASE COUNT	518 a	589 c	516 g	379 t	ORIGIN
Alignment Scores:					
Pred. No.:	3.55e-77	Length:	2002		
Score:	100.00	Matches:	100		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	67.57%	Indels:	0		
DB:	10	Gaps:	0		
US-09-809-545A-2_COPY_1_148 (1-148) x AB041596 (1-2002)					
Qy	7	ValaaenProTYThraaNGlYThRplSleuaaNProlaValaGlYAlaValaTYSeRPro	26		
Db	1237	GTCAACCCCTACCAACAAAGCTGCGAAATTAATCAGTGTGGGGCGGTCTACAGCCCC	1296		
Qy	27	AapPheTYraaGlYThrValleuLeuCYsglnAlaangIngluGlYSeRSeRMeTYr	46		
Db	1297	GACCTCTATGACGAGCAGCGGTCTGTGTGCCAGGCCAACAGAGGGAGATCTTCCATGTAC	1356		
Qy	47	SeRgLYProSeRSeRleuValTYThRSeRAlaMeTProGlyPheProTYProAla	66		
Db	1357	AGTGCCCCAGTCACTGTATATACCTTCCAAAGCTGCTCCCATCCGGCCCCC	1416		
Qy	67	ThraAlaAlaAlaTYArGAlYAlaAlaSeuArGAlYArGAlYArGThrValTYraa	86		
Db	1417	ACTGCTGCAGCTGCATACCGAGGGCTACCTTCGAGGCCGTGTGCGACCTGTACAAAC	1476		
Qy	87	ThRPeArGAlaAlaAlaProProBProBProIleProAlaTYrGlYValValTYrGln	106		
Db	1477	ACCTTCAGGGGTGCAGCGCCCCCGCCCAATCCGGCGTATGGCGGTGTATTACAG	1536		
RESULT 6	AF109106	3348 bp	mRNA	linear	PRI 01-MAR-2002
LOCUS	AF109106				
DEFINITION					Homo sapiens hexaribonucleotide binding protein 1 isoform beta
ACCESSION	AF109106				
VERSION	AF109106.1	GI:19032367			
KEYWORDS					
SOURCE					
ORGANISM					Homo sapiens.
REFERENCE					Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS					Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE					1 (bases 1 to 3348)
					Chen, W., Chu, Z.-L., Blough, R.I., Liu, L., Hoppes, B. and
					Winkelmann, J.C.
					Molecular Cloning and Chromosomal Localization of a Novel Human
					Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene
					Homologous to fox-1 in Caenorhabditis elegans
					Unpublished
					2 (bases 1 to 3348)
					Chen, W., Chu, Z.-L., Blough, R.I., Liu, L., Hoppes, B. and
					Winkelmann, J.C.
					Direct Submission
					Submitted (24-NOV-1998) Internal Medicine/Hematology-Oncology,
					University of Cincinnati College of Medicine, 231 Bethesda Ave,
					0508, Cincinnati, OH 45267-0508, USA
					Location/Qualifiers
					1. 3348
					/organism="Homo sapiens"
					/db_xref="taxon:9606"
					/chromosome="16"
					/map="16p13.3"
					1. 3348
					/gene="HRNBP1"
					254. 1441
					/gene="HRNBP1"


```
/note="similar to Caenorhabditis elegans fox-1"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AA183406.1"
/db_xref="GI:19032368"
/translation="MLASQGVLLHPYGVPMIVPAAPYLPGLQGNQEAAPPTMAOP
VSAQAPPNQNGI PAEYTAAPHAPPEYGTQTVPEHTLNLVPAQTHSEQSPADTSA
QVSGTATQTDAAPTDGPQOTQSEVENTESKSPKRLHVSNI PPRFDPDLRQMFQCF
GKILDEVI IENERGSKGFVTPENSADBARREKLGITVEGRKIEVNNAITAVMNTN
KTVNPTNCKMLNPVYGAYSPFVAGTLLCOANQSGSSWYSAPSLIVYTSAMGF
PYPAATRAAAYRGCAHLRGRTVYNTFRRAAPPPIPAVGGVYQEPYVGNKLQGGTA
AATVAPQTPATAAAYSDRNQFVVADEISNTSAVTDEFVLPPTTHLLQPPPTA
LVP"
polya_signal
3299..3304
/gene="HNRBP1"
BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN

Alignment Scores:
Pred. No.: 1,71e-67 Length: 3348
Score: 89.00 Matches: 89
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 60.14% Indels: 0
DB: 9 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x AF109106 (1-3348)
QY 49 ProSerSerLeuValTYrThrSerAlaMetProGlyPheProTYrProAlaAlaThrAla 68
Db 1043 CCCAGTTCACCTGTATATCTTGCAATGCGAGGCTTCCGTAATCCAGCAGCAGCC 1102
QY 69 AlAlaAlaAlaTYrArgGlyAlaHisLeuAlaArgGlyAlaArgGlyAlaTYrAsnThrPhe 88
Db 1103 GCGGCCCCCTACCCAGGCGCCACCTCGAGGCGCGGCTCCACCGGTACAAACACTTC 1162
QY 89 ArgAlaAlaAlaProProProProGlyLeProAlaTYrGlyGlyAlaValTYrGlnGluPro 108
Db 1163 AGGCGCGCGGCGCCCGCCCGCCATCCGCGCTACGCGGAGNAGTATACAGAGGCT 1222
QY 109 ValTYrGlyAsnLysLeuLeuGlnGlyGlyTYrAlaAlaTYrArgTYrAlaGlnProThr 128
Db 1223 GTGATGCGCAATAATGCTCAGGAGTGTATGCTGATACCGCTACGCCACGCTTACC 1282
QY 129 ProAlaThrAlaAlaAlaTYrSerAsp 137
Db 1283 CCTGCCACTGCCGCTGCTACAGTGAC 1309

RESULT 7
AF191501 1586 bp mRNA linear ROD 01-MAR-2002
LOCUS AF191501 hexaribonucleotide binding protein 1 (Hrbp1) mRNA,
DEFINITION complete cds.
ACCESSION AF191501
VERSION AF191501.1 GI:19032413
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1586)
Chen, W. and Winkelman, J. C.
Direct Submission
Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
FEATURES
Location/Qualifiers
1..1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
1..1586
gene
```

```
CDS
/gene="Hrbp1"
166..1419
/gene="Hrbp1"
/codon_start=1
/product="hexaribonucleotide binding protein 1"
/protein_id="AA183425.1"
/db_xref="GI:19032414"
/translation="MLASQGVLLHSHYGVPMIVPAAPYLPGLQGNQEAAPPTMAOP
VSAQAPPNQNGI PAEYTAAPHAPPEYGTQTVPEHTLNLVPAQTHSEQSPADTSAO
QVSGTATQTDAAPTDGPQOTQSEVENTESKSPKRLHVSNI PPRFDPDLRQMFQCF
GKILDEVI IENERGSKGFVTPENSADBARREKLGITVEGRKIEVNNAITAVMNTN
KTVNPTNCKMLNPVYGAYSPFVAGTLLCOANQSGSSWYSAPSLIVYTSAMGF
PYPAATRAAAYRGCAHLRGRTVYNTFRRAAPPPIPAVGGVYQEPYVGNKLQGGTA
AATVAPQTPATAAAYSDSYGRVYAADPYHHTLAPAPTYGVGANNAPAPLTDAKTRSHA
DVGVLVLSLQASLYRGYNRPAPY"
BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN

Alignment Scores:
Pred. No.: 1.13e-42 Length: 1586
Score: 60.00 Matches: 139
Percent Similarity: 95.86% Conservative: 0
Best Local Similarity: 95.86% Mismatches: 3
Query Match: 40.54% Indels: 6
DB: 10 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x AF191501 (1-1586)
QY 7 ValAsnProTYrThrArgGlyTYrLysLeuAsnProValGlyAlaValTYrSerPro 26
Db 826 GTCAACCCCTACACCAATGCTGGAATTAATCAATGTGGCGCGGTACAGCCCC 885
QY 27 AspPheTYrAlaGlyThrValHisLeuLeuGlnGlnAlaAsnGlnGlySerSerMetTYr 46
Db 886 GACTTCTATGCAAGCAAGCTCTGTTGTCAGGCCAACAAGAGGATCTTCATGTAC 945
QY 47 SerGlyProSerSerLeuValTYrThrSerAlaMetProGlyPheProTYrProAlaAla 66
Db 946 AGTGCCCGCAATTCATGTATATCTTGCAATGCTGAGCTTCCCATATCCGCGGCC 1005
QY 67 ThrAlaAlaAlaAlaTYrArgGlyAlaHisLeuAlaArgGlyAlaArgGlyAlaTYrVal 86
Db 1006 TC-TGCTGCAAGCTGCAACGAGGAGCTCACCTTCGAGCGCGTGCACCGG-GTACA 1063
QY 86 snThrPhe-ArgAlaAlaAlaProProProProGlyLeProAlaTYrGlyGlyAlaValTYr 105
Db 1064 ACACCTT-CAGGCTGCAAGCGCCCGCCCGCCATCCGCGCTATGCGGAGTGTAT 1122
QY 106 GlnGluProValTYrGlyAsnLysLeuLeuGlnGlyGlyTYrAlaAlaTYrArgTYrAla 125
Db 1123 CAAGAGCCAGTGTATGCGCAATAATGCTCAGGAGTGTATGCTGCGTACCGCTATGCC 1182
QY 126 GlnProThrProAlaThrAlaAlaAlaTYrSerAspSerTYrGlyAlaGlyAlaAla 145
Db 1183 CAGGCCAACCCCTGCGCTGCGCTGCTACAGTACAGTACAGCAGATTATGCTGCC 1242
QY 146 AspProTYr 148
Db 1243 GACCCCTAC 1251

RESULT 8
AB060859 1475 bp mRNA linear PRI 13-JUN-2001
LOCUS AB060859 Macaca fascicularis brain cDNA clone:QtrA-11594, full insert
DEFINITION Macaca fascicularis brain cDNA clone:QtrA-11594, full insert
ACCESSION AB060859.1 GI:13874510
VERSION AB060859.1
KEYWORDS oligo capping; file (full insert sequence).
SOURCE Macaca fascicularis adult male temporal lobe right cDNA to mRNA,
clone 1lb:macaque brain cDNA library QtrA clone:QtrA-11594.
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
```

Cercopithecinae; Macaca.

REFERENCE
AUTHORS Ooada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
TITLE Isolation of full-length cDNA clones from macaque brain CDNA
libraries

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 1475)

TITLE Hashimoto,K., Ooada,N., Hida,M., Kusuda,J. and Sugano,S.
JOURNAL Direct Submission

COMMENT Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail:khashim@nig.go.jp URL:http://www.nih.go.jp/yoken/genebank/,
Tel:-81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host: TOPI0
Vector: pMEBIS-FL3 (acc.NO. AB009864)
R. Site1: DraIII (CAGCTGTG)
R. Site2: DraII (CAGCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGTCGCCCTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraII sites of pMEBIS-FL3. XhoI sites just outside
the draII sites can be used to isolate the cDNA insert. Libraries
were constructed by oligo-capping method
(Sugano et al., , Institute of Medical Science, University of
Tokyo).
Custom primer used for sequencing
(5' end primer [CTTGCTGCTCAAGGCTGG]
3' end primer [CGACCTGCAGCTCCAGCAC]).
Location/Qualifiers
1..1475
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="NC_11594"
/sex="male"
/tissue_type="temporal lobe right"
/clone_lib="macaque brain CDNA library Qcra"
/dev_stage="adult"
90..1220
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB46877.1"
/db_xref="GI:13874511"
/translation="MEEKSGRMVQOGEAAADPTMAOPYASAFAPPNONGIPAEYVT
APHBPABEYGGOTTPHNTLNPAPDTHESPADNQTVSIGNAQTDDAPTGG
QPOPTSPENSKKOPKRLHSNI PFRPDDLLOMEGOFGLIDVELI ENERSKGK
GFVPFENSADADRAREKHGTVNSGRKLEVNNAIRARWNTKNTPNTNGKLNPVYG
AVSBPEFYAGTVLCQAQOESSMSASLSLVTSAMGFPPYPAATAAIRGHHLHG
RGRIYVFRFAAPPPIPAYGVVYODGFADIGGYAIVRYAQPFPAYVAAYSVRP
NQPFVVADEISCNTSAVTDEFMLPTPTTHLLOPPPALVP"

BASE COUNT 369 a 414 c 363 g 309 t

ORIGIN

Alignment Scores:

Pred. No.: 5.74e-41 Length: 1475

Score: 58.00 Matches: 58

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 39.19% Indels: 0

Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x AB060859 (1-1475)

OY ProSeSerLeuValTyTrHisSerAlaMetProGlyPheProTyPProAlaAlaThrAla 68

DB CCCAGTTCACCTGTAATATCTTCGCAAGCCAGGCTCCCTATCCAGCACCCGCC 884

OY AAlaAlaTyArgGlyAlaHisLeuATgGlyVArgGlyVArgThValTyAranthrPhe 88

DB GGGGGCCGCTACCAGAGGGCGGACCTGCGAGGCGGGGGCGGACCGTGTACAACACTTC 944

Oy	89	ArgAlaAlaAlaProProProPheIleSerAlaTyrGlyGlyValAlaTyrGln	106
Db	945	AGGGCCGACGGCCCCCCCCCCCATCCCCGGCTATGGCGGTGTTGTTTACCAG	998
RESULT 9			
AK001027		1513 bp	linear
LOCUS			PRI 01-AUG-2002
DEFINITION			Homo sapiens cDNA FLJ10165 fis, clone HEMBA1003591, weakly similar to CHIRODIPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
COMMENT			
FEATURES			
Source			
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best local Similarity:			
Query Match:			
DB:			

US-09-809-545a-2_COPY_1_148 (1-148) x AK001027 (1-1513)

Qy 49 ProSerSerLeuValIYrThrSerAlaMetProGlyPheProIYrProAlaAlaThrAla 68
Db 950 CCGAGTTCACCTTGATATACCTTCTGCAATGCCAGGCTCCGATCCAGCAGCCAGCC 1009

Qy 69 A1a1a1a1a1YrArgIyAlaHisLeuArgIyArgIyArgThrValIYrAsnThrPhe 88
Db 1010 GCGGCGGCTACCGAGGGGCGACCTCGAGGCGCGGTCGACCGGTATACACACCTTC 1069

Qy 89 ArgAlaAlaAlaProProProProIleProAlaIYrGlyIyValIYrGln 106
Db 1070 AGGCGCGGCGGCGCCCGCCCGATCCCGCTACGCGGCGGTGTATTACAG 1123

RESULT 10
AF107203

LOCUS AF107203 2279 bp mRNA linear PRI 23-JUN-2000
DEFINITION Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.
ACCESSION AF107203
VERSION AF107203.1 GI:8671585
KEYWORDS

SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2279)
Shibata, H., Huynh, D. P., Vo, T. T. and Pulst, S.-M.
A novel protein, A2BP, with RNA binding motif binds to C-terminal
ataxin-2
Unpublished
2 (bases 1 to 2279)
Shibata, H., Huynh, D. P., Vo, T. T. and Pulst, S.-M.
Direct Submission
Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA
Location/Qualifiers
1..2279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="similar to Homo sapiens cDNA clone HSP38A20"

FEATURES
source 1..2279
/gene="A2BP"
1047..2180
/gene="A2BP"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AF78291.1"
/db_xref="GI:8671586"
/translation="MAQPIYASQAPAPONGIPATYTPAHPHAPETYGQTTVPENTLN
LYPPAQTHSESPADTSAGTQVATOTDDAPDGPQVPSNTENKSOQKSLHVS
NIPRFRPDRLQMFQFGKILDEIIFENSGSGFGFVFEENADRAEKLAGTV
VEGKRIEIVNNAATNTKNTVNPYNTGMKYNPVGAVYSEFYAGTVLLCOAQEES
SMYGAPSLSLYTSAMPGPYPYAPAAAYAKYNAHGRGRIVYTFRRAAPPPIDAG
GVVGDGFTYGADIVGYAAYRYAOPTATAAAYSDYGRVYADPHYHALAPATYGV
GAMNAFPIIDAKTRSHADVGLVLSLQASIVYGNRPAPY"

misc_feature 1341..1358
/gene="A2BP"
/note="Region: RNA binding motif RNP-2"
misc_feature 1452..1472
/gene="A2BP"
/note="Region: RNA binding motif RNP-1"

BASE COUNT 497 a 712 c 636 g 434 t
ORIGIN

Alignment Scores:
Pred. No.: 8.03e-41 Length: 2279
Score: 58.00 Matches: 58
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.19% Indels: 0

DB: 9 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x AF107203 (1-2279)

Qy 49 ProSerSerLeuValIYrThrSerAlaMetProGlyPheProIYrProAlaAlaThrAla 68
Db 1716 CCGAGTTCACCTTGATATACCTTCTGCAATGCCAGGCTCCGATCCAGCAGCCAGCC 1775

Qy 69 A1a1a1a1a1YrArgIyAlaHisLeuArgIyArgIyArgThrValIYrAsnThrPhe 88
Db 1776 GCGGCGGCTACCGAGGGGCGACCTCGAGGCGCGGTCGACCGGTATACACACCTTC 1835

Qy 89 ArgAlaAlaAlaProProProProIleProAlaIYrGlyIyValIYrGln 106
Db 1836 AGGCGCGGCGGCGCCCGCCCGATCCCGCTACGCGGCGGTGTATTACAG 1889

RESULT 11
AR134676

LOCUS AR134676 2372 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6194171.
ACCESSION AR134676
VERSION AR134676.1 GI:14123581
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
1 (bases 1 to 2372)
Pulst, S.M. and Shibata, H.
Nucleic acids encoding ataxin-2 binding proteins
Patent: US 6194171-A 1 27-FEB-2001;
Location/Qualifiers
1..2372
/organism="unknown"

BASE COUNT 548 a 726 c 646 g 452 t
ORIGIN

Alignment Scores:
Pred. No.: 6.06e-40 Length: 2372
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.51% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x AR134676 (1-2372)

Qy 49 ProSerSerLeuValIYrThrSerAlaMetProGlyPheProIYrProAlaAlaThrAla 68
Db 1716 CCGAGTTCACCTTGATATACCTTCTGCAATGCCAGGCTCCGATCCAGCAGCCAGCC 1775

Qy 69 A1a1a1a1a1YrArgIyAlaHisLeuArgIyArgIyArgThrValIYrAsnThrPhe 88
Db 1776 GCGGCGGCTACCGAGGGGCGACCTCGAGGCGCGGTCGACCGGTATACACACCTTC 1835

Qy 89 ArgAlaAlaAlaProProProProIleProAlaIYrGlyIyValIYrGln 105
Db 1836 AGGCGCGGCGGCGCCCGCCCGATCCCGCTACGCGGCGGTGTATTAC 1886

RESULT 12
HSFOX09

LOCUS HSFOX09 251 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 (HRNB1) gene,
exon 9.
ACCESSION AF109115
VERSION AF109115.1 GI:19032378
KEYWORDS

SOURCE 9 of 14
Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 251)
Chen, W. and Winkelman, J. C.


```

repeat_region /rpt_family="L1"
43924. .44555
repeat_region /rpt_family="L1"
45128. .45367
repeat_region /rpt_family="Alu"
45467. .45704
repeat_region /rpt_family="Alu"
46192. .46578
repeat_region /rpt_family="MER25"
complement(46779. .47141)
repeat_region /rpt_family="THE1"
complement(48265. .48550)
repeat_region /rpt_family="Alu"
49630. .50221
repeat_region /rpt_family="MER41"
50781. .51096
repeat_region /rpt_family="Alu"
complement(51219. .51439)
repeat_region /rpt_family="Alu"
51462. .51708
repeat_region /rpt_family="MER33"
complement(52042. .52321)
repeat_region /rpt_family="Alu"
54001. .54092
repeat_region /rpt_family="Alu"
complement(55760. .56014)
repeat_region /rpt_family="Alu"
56616. .56899
repeat_region /rpt_family="Alu"
complement(66463. .66747)
repeat_region /rpt_family="Alu"
complement(66908. .67180)
repeat_region /rpt_family="Alu"
complement(70389. .70704)
repeat_region /rpt_family="MER1"
complement(72166. .72421)
repeat_region /rpt_family="MER33"
72501. .72769
repeat_region /rpt_family="Alu"
72643. .72956
repeat_region /rpt_family="MER5"
complement(76503. .76564)
repeat_region /rpt_family="Alu"
97534. .97700
repeat_region /rpt_family="MER3"
complement(98469. .98526)
repeat_region /rpt_family="Alu"
complement(99089. .99381)
repeat_region /rpt_family="Alu"
101648. .101930
repeat_region /rpt_family="Alu"
102790. .103056
repeat_region /rpt_family="Alu"
104010. .104277
repeat_region /rpt_family="Alu"
complement(111121. .111263)
repeat_region /rpt_family="MER5"
complement(112263. .112476)
repeat_region /rpt_family="MER30"
112956. .113234
repeat_region /rpt_family="Alu"
114385. .114657
repeat_region /rpt_family="Alu"
complement(117031. .117194)
repeat_region /rpt_family="MER20"
complement(118464. .118770)
repeat_region /rpt_family="Alu"
119271. .119549
repeat_region /rpt_family="Alu"
complement(122831. .122940)
repeat_region /rpt_family="MER41"
123471. .123742
repeat_region /rpt_family="Alu"

```

```

repeat_region complement(129860. .129972)
repeat_region /rpt_family="Alu"
135235. .135310
repeat_region /rpt_family="MLT1"
135874. .136157
repeat_region /rpt_family="Alu"
139775. .140048
repeat_region /rpt_family="Alu"
141734. .141814
repeat_region /rpt_family="MIR"
143071. .143352
repeat_region /rpt_family="Alu"
148286. .148773
repeat_region /rpt_family="MER1"
150431. .150707
repeat_region /rpt_family="Alu"
155151. .155387
repeat_region /rpt_family="Alu"
complement(156101. .156213)
repeat_region /rpt_family="MER20"
complement(156636. .156920)
repeat_region /rpt_family="Alu"
complement(159633. .159931)
repeat_region /rpt_family="Alu"
complement(160378. .160658)
repeat_region /rpt_family="Alu"
161617. .162140
repeat_region /rpt_family="Alu"
complement(163200. .163315)
repeat_region /rpt_family="MIR"
BASE COUNT 48114 a 34256 c 33738 g 48430 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.77e-27 Length: 164538
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.73% Indels: 0
DB: 9 Gaps: 0

```

US-09-809-545a-2_COPY_1_148 (1-148) x ACC005774 (1-164538)

```

QY 59 ProGlyPheProTYrPCoAlaAlaThrAlaAlaAlaTYrAGGlyAlaHisLeuArg 78
Db 118155 CAGGCTTCCCGTATCCAGCGACCGCCGCCGCTACCGAGGGCCACCTGGCA 118096
QY 79 G1yArG1yArGThrValTYrAsnThrPheArGAlaAlaAlaProProPro11ePro 98
Db 118095 GGCGCGGTCGACCGTGTACACACCTTCAGGCGCCGCCGCCGCCCGCGATCCCG 118036
QY 99 AlATyrg1yGly1y 102
Db 118035 GCCTACGGCGGG 118024

```

```

RESULT 14
HSA323688 596 bp DNA linear PRI 18-JUL-2002
LOCUS HSA323688
DEFINITION Homo sapiens genomic sequence surrounding Not1 site, clone
ACCESSION N66-793C.
VERSION AJ323688
KEYWORDS AJ323688.1 GI:15868067
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 596)
AUTHORS Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvaasha,S.M.,
Podowski,R.M., Matshushin,Y.G., Gyanchandani,A., Muravenko,O.V.,
Levitsky,V.G., Kolchanov,N.A., Protodopov,A.I., Kashuba,V.I.,
Kiselev,L.L., Wasserman,W., Wallestedt,C. and Zabarovsky,E.R.

```

TITLE Not1 flanking sequences: a tool for gene discovery and verification

```
Search completed: March 16, 2003, 04:36:47
Job time : 1731 secs
```


XX	graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
KM	Alzheimer's disease; gene therapy.
XX	
OS	Rattus norvegicus.
XX	
PN	WO200174901-A2.
XX	
PD	11-OCT-2001.
XX	
PF	23-MAR-2001; 2001WO-US09555.
XX	
PR	31-MAR-2000; 2000US-193548P.
PR	14-MAR-2001; 2001US-0809545.
XX	
PA	(SCIO-) SCIOS INC.
XX	
PI	Stanton LM, White RT;
XX	
DR	WPI; 2002-010779/01.
DR	P-PsDB; AAU70146.
XX	
PT	Novel secreted factor polypeptide useful for treating cardiac diseases
PT	such as arteriosclerosis, myocardial infarction, inflammatory diseases
XX	such as asthma, stroke, and rheumatoid arthritis and renal diseases -
XX	
SS	Claim 1; Fig 1; 189pp; English.

CC The invention relates to rat secreted factor polypeptides and the
CC polynucleotides encoding them. The sequences are useful for treating
CC cardiac, renal or inflammatory diseases. These include cardiac diseases
CC such as congestive heart failure, myocarditis, dilated congestive
CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
CC arrhythmia, pulmonary hypertension, atherosclerosis and
CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
CC disease, chronic renal failure, renal vein thrombosis and medullary
CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
CC disease. Sequences MAS94693-AM94975 represent cDNA clones, which encode
CC the secreted factor polypeptides of the invention, and oligonucleotide
CC probes and PCR primers.

Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;

Size of the Dataset:	2,346,127	Length:	1340
Pred. No.:	148,00	Matches:	148
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	24	Gaps:	0
DB:	24		

US-09-809-545A-2_COPY_1_148 (1-148) X AAS94693 (1-1340)

[illegible]

Db	775	GGTGGCACCCTGATCAACACCTTCAGAGCTGGGGCGCCCAACCCCAATCCGGGCTAT	834
Qy	101	GIYGLYVAIVAIYTYGINGIUPROVALITYRGLYASNLYSLEULENGINGIYGLYTYTALA	120
Db	835	GGCCGAGTAGTGTATCAAGAGCCAGTGTATAGCATAAATTTGCTACAGGGGTGTTAGCT	894
Qy	121	ALATYRAGTYRZLAAGINPOTHPRVALATHRALAIAIATYRSEASPSETYRGLY	140
Db	895	GCAATCCGCTACGCCCGACCCCTGCCACCTGCTGCTCTCAACATGACAGTTACGGA	954
Qy	141	ARGVALTYRZLAIAASPPTYR	148
Db	955	CGAGTTATGCTGCCGACCCCTTAC	978

RESULT 2
AAH13824

ID AAH13824 standard; cDNA; 1513 BP.

AC AAH13824;

DT 26-JUN-2001 (first entry)

DE	Human cDNA sequence SEQ ID NO:10786.
DE	Human cDNA sequence SEQ ID NO:10786.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126

PR 29-JUL-1999; 99JP-0248036

PR 11-JAN-2000; 2000JP-0118776

PR 09-JUN-2000; 2000JP-0241899

PA (HELI -) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,

XX 3001 310740/34

XX
XX

full-length cDNAs defined in the specification, and for the detection

PT full-length cDNAs -

PS Claim 8; SEQ ID 10786; 2537pp + CD ROM; English

CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;
Alignment Scores:
Pred. No.: 3.57e-44 Length: 1513
Score: 58.00 Matches: 58
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.19% Indels: 0
DB: 22 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x AAH13824 (1-1513)
QY 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
DB 950 CCCAGTTCACCTGTATATACCTTCTGCAATGCCAGGCTTCCCGTATCCAGCCACCGCC 1009
QY 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyValTyrAsnThrPhe 88
DB 1010 GCGGCGGCTACCGAGGGCGCACCTGCGAGGCCGCGGTCCGACCGGTACACCTTC 1069
QY 89 ArgAlaAlaAlaProProProProPheProAlaTyrGlyGlyValValTyrGln 106
DB 1070 AGGGCGGCGGGCCCCCGCCCGATCCGGGCTACGGGGGTGTGTTACAG 1123
RESULT 3
AAK52245
ID AAK52245 standard; cDNA; 1800 BP.
XX
AC AAK52245;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 790.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhtman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79112.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 2643-2645; 6221pp; English.

XX
CC The invention relates to polynucleotides (AAK51456-AAK51435) and the
CC encoded polypeptides (AAM78333-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
Alignment Scores:
Pred. No.: 4.18e-44 Length: 1800
Score: 58.00 Matches: 58
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.19% Indels: 0
DB: 22 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x AAK52245 (1-1800)
QY 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
DB 1237 CCCAGTTCACCTGTATATACCTTCTGCAATGCCAGGCTTCCCGTATCCAGCCACCGCC 1296
QY 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyValTyrAsnThrPhe 88
DB 1297 GCGGCGGCTACCGAGGGCGCACCTGCGAGGCCGCGGTCCGACCGGTACACACCTTC 1356
QY 89 ArgAlaAlaAlaProProProProPheProAlaTyrGlyGlyValValTyrGln 106
DB 1357 AGGGCGGCGGGCCCCCGCCCGATCCGGGCTACGGGGGTGTGTTACAG 1410
RESULT 4
AAA07075
ID AAA07075 standard; cDNA; 2372 BP.
XX
AC AAA07075;
XX
DT 03-JUL-2000 (first entry)
XX
DE cDNA encoding human ataxin-2 binding protein (A2BP).
XX
KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
KW neuronal plasticity; cellular degeneration signal transduction pathway;
KW selective RNA transport; spinocerebellar ataxia type-2;
KW hyperproliferative disorder; ss.
XX
OS Homo sapiens.
XX
PN WO2000012710-A1.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20156.
XX
PR 01-SEP-1998; 98US-0145391.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Pulst SM, Shibata H;

XX WPI, 2000-237873/20.
 DR P-PSDB; AAY81462.
 XX Nucleic acids encoding an ataxin-2 binding protein useful for
 PT inhibiting the expression of active proteins from the SCA2 gene for the
 PT treatment of spinocerebellar ataxia type-2.
 XX
 XX Claim 6; Page 74-77; 82pp; English.
 PS
 CC This sequence represents cDNA encoding human ataxin-2 binding protein
 CC (A2BP). Nucleotide sequences encoding human A2BP were originally
 CC isolated in an adult brain cDNA library using the yeast two hybrid
 CC method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
 CC function that is encoded by the SCA2 gene located on chromosome 12. SCA2
 CC has been linked to the autosomal dominant neurodegenerative disorder
 CC spinocerebellar ataxia type-2. Individuals afflicted with the disease
 CC exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
 CC containing a polyglutamine stretch of about 35-39 residues, whereas that
 CC of normal individuals contains approximately 22 contiguous glutamine
 CC residues. A2BP and ataxin-2 are components of a cellular degeneration
 CC signal transduction pathway. The pathogenic expanded form of ataxin-2 has
 CC a higher affinity for A2BP relative to normal ataxin-2; the presence of
 CC the expanded form is likely to promote degeneration. A2BP and ataxin have
 CC also been found to have a role in gene regulation. The binding of A2BP to
 CC ataxin-2 plays an important role in controlling gene expression via the
 CC targeting of transport of specific RNAs, selective RNA transport being
 CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
 CC early in embryonic development. Both ataxin-2 and A2BP are able to bind
 CC RNA, and are essential components of the RNA localization network that
 CC establishes cellular polarity in embryogenesis. In highly differentiated,
 CC polarized cells such as neurons, A2BP and ataxin-2 have a similar
 CC function and are required for neuronal plasticity. A2BP nucleic acids may
 CC be used for the recombinant production of A2BP proteins or fragments
 CC thereof according to standard methodologies. For example, an A2BP protein
 CC with an ataxin-2 or RNA binding capability but no signal transduction
 CC function can be used as a dominant negative inhibitor of the cellular
 CC degeneration signal transduction pathway. A2BP proteins with a signal
 CC transduction function can be used to treat hyperproliferative disorders
 CC (e.g., cancer) via stimulation of the cellular degeneration pathway.
 XX
 SQ Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,52e-43 Length: 2372
 Score: 57.00 Matches: 57
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38,514 Indels: 0
 DB: 21 Gaps: 0
 US-09-809-545A-2_COPY_1_148 (1-148) x AAA07075 (1-2372)
 QY 49 ProSerSetLeuValTyrrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
 DB 1716 CCGAGTTCACTGTATATCTTCTGCATGCGAGCTTCCCGATATCAGAGCACCGCC 1775
 QY 69 AAlaAlaAlaTyrArgGlyValAlaHisLeuArgGlyAArgGlyAArgThrValTyrAanThrPhe 88
 DB 1776 GGGGGCGCTTACCGAGGGGGGACCTGCGAGGGCGGCTGCGACCGTGTACACACCTTTC 1835
 QY 89 ArgAlaAlaAlaProProProProProIleProAlaTyrGlyValValTyr 105
 DB 1836 AGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 1886
 RESULT 5
 AAS93634
 ID AAS93634 standard; cDNA; 2118 BP.
 XX AAS93634;
 AC
 XX 13-FEB-2002 (first entry)
 DT
 XX

DE DNA encoding novel human diagnostic protein #29438.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX Food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI, 2001-639362/73.
 DR P-PSDB; ABG29447.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1, SEQ ID No 29438; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensic, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,15e-31 Length: 2118
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 29,73% Indels: 0
 DB: 23 Gaps: 0
 US-09-809-545A-2_COPY_1_148 (1-148) x AAS93634 (1-2118)
 QY 59 ProGlyPheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyValAlaHisLeuArg 78
 DB 586 CCGAGCTTCCCGTATCCAGACCCCGCGGGCGGCTTACCGAGGGGGGACCTGCGCA 645
 QY 79 GlyArgGlyAArgThrValTyrAsnThrPheArgAlaAlaAlaProProProIlePro 98
 DB 646 GGGCGGGCTGCGACCGTGTACACACCTTCAAGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 705
 QY 99 AlaTyrGlyGly 102

Db 706 GCCTACCGCGGT 717
 RESULT 6
 AAS93633
 ID AAS93633 standard; cDNA; 481 BP.
 AC AAS93633;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29437.
 XX
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG28446.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 29437; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS6197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,73e-26 Length: 481
 Score: 38.00 Matches: 38
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.68% Indels: 0
 DB: 23 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x AAS93633 (1-481)
 QY 69 AlaAlaAlaArgGlyAlaHisLeuArgGlyArgGlyArgThrValTYrAsnThrPhe 88
 Db 1 GCGGCCCGCTTACCGAGGGGGGCACTTCGAGCGCGGTGCACGGTACACACCTTC 60
 QY 89 ArgAlaAlaAlaProProProProIleProAlaTYrGlyValValTYrGln 106
 Db 61 AGGCGCGCGCGCGCGCGCGCGCGATCCGCTACGCGCGGTGTATTACAG 114
 RESULT 7
 ID ABR34530 standard; cDNA; 539 BP.
 AC ABR34530;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human cDNA for novel secreted protein, SEQ ID 299.
 XX
 KM Human; ss; gene; secreted protein; immune deficiency; viral infection;
 KM bacterial infection; fungal infection; autoimmune disorder; burn;
 KM rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
 KM diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
 KM Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
 KM coagulation disorder; haemophilia; inflammatory disorder; ulcer;
 KM tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
 KM lymphoid cell deficiency.
 XX
 OS Homo sapiens.
 XX
 PN WO200177290-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 29-MAR-2001; 2001WO-US10295.
 XX
 PR 06-APR-2000; 2000US-194941P.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Wong GG, Clark HF, Fecthel K, Agostino MJ, Howes SH, Resnick RJ;
 PI Gutukota K, Graham JR;
 XX
 DR WPI: 2002-179323/23.
 XX
 PT Six hundred and twenty five polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders
 PT -
 XX
 PS Claim 1; Page 153; 339pp; English.
 XX
 CC The invention relates to 625 polynucleotides which have been derived from
 CC a variety of human tissue sources and sequences that hybridize to them.
 CC proteins, their complements and sequences that hybridize to them.
 CC Also included are a vector comprising the polynucleotide, a host cell
 CC transformed with the vector, the proteins encoded by the
 CC polynucleotides, antibodies that bind to the proteins and identification
 CC of modulators of the proteins or the expression of the polynucleotide.
 CC The polynucleotides can be used as probes for the identification
 CC and isolation of full length cDNA and genomic DNA. The polynucleotides
 CC and proteins can also be used as nutritional supplements. The protein
 CC is useful in the treatment of various immune deficiencies and disorders
 CC such as viral infections, bacterial infections, fungal infections,
 CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
 CC autoimmune thyroiditis and diabetes) and allergic reactions and
 CC conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment
 CC of burns, incisions and ulcers. The proteins are also useful for

CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.
XX
SQ Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;
Alignment Scores:
Pred. No.: 2,44e-23 Length: 539
Score: 35.00 Matches: 35
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.65% Indels: 0
DB: 24 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x ABK34530 (1-539)
QY 72 TYRARGGLYALHIELEUARGGLYARGGLYRVAITYRFAHTRHPhEARGLAa1a1a 91
DB 538 TACCGAGGGGCGACCTGGAGGCGCGGTCGACCGTACAAACCTTCAGGGCGCG 479
QY 92 A1AProProProProleProalaTYrGLYGLYValValTYrGln 106
DB 478 GGGCCCCCCCCCGATCCCGGCTACGGCGGTGTTCACAG 434
RESULT 8
AAH09205/C
ID AAH09205 standard; cDNA; 607 BP.
XX AC AAH09205;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (3'-primer) SEQ ID NO:6040.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX PS Claim 3; SEQ ID 6040; 2537bp + CD ROM; English.
XX CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 607 BP; 140 A; 152 C; 179 G; 129 T; 7 other;
Alignment Scores:
Pred. No.: 1.61e-20 Length: 607
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.62% Indels: 0
DB: 22 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x AAH09205 (1-607)
QY 117 GLYGLYTYRAlaAlaTYrAlaTYrAlaGlnProThrProAlaThrAlaAlaTYrSer 136
DB 363 GGTGGTATGCTGCAATACCGCTACGCCGCTACCCCTGCGACATCGCGCTCAACT 304
QY 137 AepSerTYrGLYARGValTYrAlaAlaAspProTYr 148
DB 303 GACAGTTACGAGCAGATTATGCTCGCAGCCCTAC 268
RESULT 9
AAK53229/C
ID AAK53229 standard; cDNA; 1164 BP.
XX AC AAK53229;
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 2758.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YR, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX DR WPI; 2001-476283/51.

DR P-PSDB; AAM80096.
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 XX Claim 1; Page 4968, 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.54e-09 Length: 1164
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservatave: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 13.51% Indels: 0
 DB: 22 Gaps: 0
 US-09-809-545a-2_COPY_1_148 (1-148) x AAK53229 (1-1164)
 Qy 7 ValaenProTYrThAsnGlyTrpLysLeuAnProValaIglYalaIvalTYrSerPro 26
 Db 254 GTCAACCTTATACAAATGGCTGGAATTCAGTTGTGGTGCGAGTCTACAGTCCC 195
 RESULT 10
 AAS66245/C
 ID AAS66245 standard; cDNA; 567 BP.
 XX
 AC AAS66245;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #2049.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG02058.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX Claim 1; SEQ ID No 2049; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (II) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 567 BP; 92 A; 151 C; 169 G; 155 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.08e-06 Length: 567
 Score: 17.00 Matches: 17
 Percent Similarity: 100.00% Conservatave: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 11.49% Indels: 0
 DB: 23 Gaps: 0
 US-09-809-545a-2_COPY_1_148 (1-148) x AAS66245 (1-567)
 Qy 8 AsnProTYrThAsnGlyTrpLysLeuAnProValaIglYalaIvalTYr 24
 Db 480 AACCCCTACACCAACGCGCTGAGAGCTAATCAGTGGTGGCGGAGTCTAC 430
 RESULT 11
 AAI59249
 ID AAI59249 standard; cDNA; 918 BP.
 XX
 AC AAI59249;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1452.
 XX
 KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB; AAM40093.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 1452; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilization of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX SQ Sequence 918 BP; 235 A; 300 C; 235 G; 148 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1,68e-06 Length: 918
XX Score: 17.00 Matches: 17
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 11.49% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545a-2_COPY_1_148 (1-148) x AA159249 (1-918)
XX
XX Oy 8 AaPProTYThRAsnGlyTrrPLySLeuaSProValValGlyAlaValTyr 24
XX DB 459 AACCCCTACACCAAGCGCTGGAAGCTAAATCCAGTGTGGCGCACTTAC 509
XX
XX RESULT 12
XX AA161034/C
XX ID AA161034 standard; cDNA; 939 BP.
XX
XX AC AA161034;
XX
XX DT 22-OCT-2001 (first entry)
XX
XX DE Human polynucleotide SEQ ID NO 5023.
XX
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200153312-A1.
XX
XX PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000WO-US34263.

XX
XX 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB; AAM41878.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 5023; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilization of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.72e-06 Length: 939
XX Score: 17.00 Matches: 17
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 11.49% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545a-2_COPY_1_148 (1-148) x AA161034 (1-939)
XX
XX Oy 8 AaPProTYThRAsnGlyTrrPLySLeuaSProValValGlyAlaValTyr 24
XX DB 480 AACCCCTACACCAAGCGCTGGAAGCTAAATCCAGTGTGGCGCACTTAC 430
XX
XX RESULT 13
XX AA161035/C
XX ID AA161035 standard; cDNA; 939 BP.
XX
XX AC AA161035;
XX
XX DT 22-OCT-2001 (first entry)
XX
XX DE Human polynucleotide SEQ ID NO 5024.
XX
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.

RESULT 15
AAS90506

ID AAS90506 standard; cDNA; 578 BP.

AC AAS90506;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #26310.

KW Human, chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder; ss

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001

PF 30-MAR-2001; 2001WO-US08631

PR 31-MAR-2000; 2000US-0540217

PR 23-AUG-2000; 2000US-0649167

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT

DR WPI; 2001-639362/73.

DR P-PSDB; ABG26319.

PT New isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity -

PS Claim 1; SEQ ID No 26310; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridization probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or quantifying a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

NOTE: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

SQ Sequence 578 BP; 152 A; 165 C; 146 G; 113 T; 2 others?

Alignment Scores:

Pred. No.:	0.00546	length:	578
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.78%	Indels:	0
DB:	23	Gaps:	0

US-09-809-545A-2_COPY_1_148 (1-148) X AAS90506 (1-578)

73 ArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyr 85

Db 154 AGAGGAGCCCATTTGAGGGCGAGAGGGCGACATATAT 192

Search completed: March 16, 2003, 04:07:45
Job time : 225 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 04:00:32 ; Search time 1442 Seconds

(without alignments)
1662.227 Million cell updates/sec

Title: US-09-809-545a-2_COPY_1_148
Perfect score: 148
Sequence: 1 MTRKKAVPYTNCKMLNPV.....PATMAAYSDSYGRVADPY 148

Scoring table:

OLIGO	Xgapop 60.0 , Xgapext 60.0
	Xgapop 60.0 , Xgapext 60.0
	Xgapop 6.0 , Xgapext 7.0
	Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DB=xlp
-Q=/cgn2_1/USFTO.spool/US09809545/runat_07032003_153955_12169/app_query.fasta_1.327
-DB=EST -QMT=fastcap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco
-NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=200000000
-USR=US09809545 @CGN 1.1 2874 @runat_07032003_153955_12169 -NCPU=6 -ICPU=3
-MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1:	em_estba:*
2:	em_esthum:*
3:	em_estlin:*
4:	em_estma:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hic:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_hic:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vrt:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_yod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	54.1	391	9 A1186273	A1186273 qd20b08.x
2	80	54.1	495	9 A1656926	A1656926 t448e08.x
3	58	39.2	4039	11 BC026312	BC026312 Homo sapi
4	48	32.4	533	12 BF549922	BF549922 UI-R-E0-b
5	38	25.7	477	10 AW197589	AW197589 XM44G02.x
6	38	25.7	480	9 A1401040	A1401040 ch27a12.x
7	38	25.7	773	10 AV729198	AV729198 AV729198
8	38	25.7	774	10 AV729057	AV729057 AV729057
9	36	24.3	439	9 A1968176	A1968176 wu14a06.x
10	36	24.3	439	10 AW589795	AW589795 hg22a04.x
11	36	24.3	439	10 BE501380	BE501380 T441n09.x
12	35	23.6	261	14 F21037	F21037 HSPD05503.H
13	35	23.6	383	9 A1189407	A1189407 NB6-793C
14	34	23.0	589	17 AQ938012	AQ938012 qc99h03.x
15	32	21.6	218	9 A1189257	A1189257 qc99h03.x
16	32	21.6	367	9 A1095813	A1095813 qb20911.x
17	32	21.6	372	9 A1799929	A1799929 wc41a05.x
18	32	21.6	420	9 A1655094	A1655094 wb67a10.x
19	32	21.6	449	12 BF223478	BF223478 T33601.x
20	32	21.6	449	9 A1374891	A1374891 ta60912.x
21	32	21.6	466	9 A1917298	A1917298 t896909.x
22	32	21.6	459	9 A1244212	A1244212 q186b11.x
23	31	20.9	607	9 A1975235	A1975235 q33608.x
24	31	20.9	392	9 AA975235	AA975235 q33608.x
25	29	19.6	87	9 A1240005	A1240005 qb33606.x
26	28	18.9	408	12 BC795922	BC795922 TMSM SM38
27	28	18.9	521	13 BM183349	BM183349 t10e03.x
28	25	16.9	443	9 AA773715	AA773715 a181c05.x
29	25	16.9	733	12 BC306387	BC306387 tm86d11.x
30	23	15.5	655	14 BQ188308	BQ188308 UI-E-EJ1-
31	22	14.9	451	9 A1335996	A1335996 qf42e01.x
32	22	14.9	451	9 A1338843	A1338843 qe72h11.x
33	21	14.2	423	14 R88535	R88535 ym91912.61
34	21	14.2	553	9 AA460647	AA460647 zv72e08.6
35	20	13.5	291	14 Z25293	Z25293 HSB8B042.S
36	20	13.5	293	9 AA621734	AA621734 af54h09.s
37	20	13.5	302	14 BM930547	BM930547 UI-E-EJ1-
38	20	13.5	304	9 A1962834	A1962834 wq43f08.x
39	20	13.5	306	9 AA405518	AA405518 zw36f08.b
40	20	13.5	323	10 AW295337	AW295337 UI-H-B12-
41	20	13.5	347	9 A1041587	A1041587 ox56g05.x
42	20	13.5	428	9 AA460940	AA460940 zx61d04.s
43	20	13.5	428	10 AV665024	AV665024 AV665024
44	20	13.5	434	13 BM031289	BM031289 a46653.MA
45	20	13.5	457	9 A1193117	A1193117 qee5h03.x

ALIGNMENTS

RESULT 1
LOCUS A1186273
DEFINITION qd20b08.x1 Soares placenta Bco9week 2bhbP8tc9w Homo sapiens CDNA
clone IMAGE:1724247 3', mRNA sequence.
ACCESSION A1186273
VERSION A1186273.1 GI:3736911
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 391)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE

TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcg@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R..M.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 32 Row: 1 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922072
This clone has the following problem: frame shifted.

FEATURES
source
1. 4039
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4815500"
/tissue_type="Brain, hippocampus"
/clone_id="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"

BASE COUNT 1201 a 858 c 876 g 1104 t
ORIGIN

Alignment Scores:
Pred. No.: 2,23e-43 Length: 4039
Score: 58.00 Matches: 58
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.19% Indels: 0
DB: 11 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x BC026312 (1-4039)

Qy 49 ProSerSerLeuValTYrThSerAlaMetProGlyPheProTYrProAlaAlaThAla 68
|||
Db 967 CCGAGTTCACTGTATATACCTTCGCAATGCCAGGCTTCCGTAACGACGACCGCC 1026

Qy 69 AlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArgGlyThValTYrAsnThrPhe 88
|||
Db 1027 GCGGCGCGCTACCGAGGCGGCGACCTGCGAGGCGCGGTCGACCGTATACACACCTTC 1086

Qy 89 ArgAlaAlaAlaProProProProProAlaTYrGlyValValTYrGln 106
|||
Db 1087 AGGGCCGCGCGCGCCGCCGATCCGCGCTACGCGGCGGTGTTTACCAAG 1140

RESULT 4
BF549922/c 533 bp mRNA linear EST 12-DEC-2000
LOCUS BF549922
DEFINITION UI-R-E0-bv-c-04-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-bv-c-04-0-UI 5', mRNA sequence.
ACCESSION BF549922
VERSION BF549922.1 GI:11659652
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 533)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: meares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
Clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE ID= 1777166
Seq primer: M13 Forward
Location/Qualifiers

FEATURES
source
1. 533
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-UI"
/clone_id="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site 1: NotI; Site 2: EcoRI. This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 134 a 125 c 164 g 109 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 8.44e-35 Length: 533
Score: 48.00 Matches: 48
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.43% Indels: 0
DB: 12 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x BF549922 (1-533)

Qy 59 ProGlyPheProTYrProAlaAlaThAlaAlaTYrArgGlyAlaHisLeuArg 78
|||
Db 460 CCGGCTTCCATATCCGCGCGCCGACGCTGCATGCCAGGCGGCTCACCTTCA 401

Qy 79 GlyArgGlyArgThValTYrAsnThrPheArgAlaAlaProProProProPro 98
|||
Db 400 GCGCGTGTCCACCGGTATACACCTTCAGACTGCGGCGCCGCCACCAATCCG 341

Qy 99 AlaTYrGlyGlyValValTYrGln 106
|||
Db 340 GCCATAGCGGCGGTGTTTACCAAG 317

RESULT 5
AM197589 477 bp mRNA linear EST 29-NOV-1999
LOCUS AM197589
DEFINITION xm44602.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2687090 3' similar to TR:043251 043251 HYPOTHETICAL 39.5 KD PROTEIN. ., mRNA sequence.
ACCESSION AM197589
VERSION AM197589.1 GI:6476819
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 477)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 04:02:33 ; Search time 74 Seconds
(without alignments)
613.354 Million cell updates/sec

Title: US-09-809-545A-2_COPY_1_148
Sequence: 1 MTRKKAVPYTNMGKLNPPV.....PATMAAYSDSYGRVADPY 148

Scoring table:

	OLIGO
Xgapop 60.0 ,	Xgapext 60.0
Xgapop 60.0 ,	Xgapext 60.0
Xgapop 6.0 ,	Xgapext 7.0
Delop 6.0 ,	Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878600

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1p
-Q=/cgn2_1/USFTPO.spool/US09809545/runat.07032003.153955.12186/app.query.fasta.1.327
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=olip2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09809545 @CGN 1.1 61 @runat.07032003.153955.12186 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfltest.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	38.5	2372	4 US-09-145-391-1	Sequence 1, Appli
2	13	8.8	1506	3 US-09-176-657-5	Sequence 1, Appli
3	8	5.4	1007	3 US-09-248-335-31	Sequence 31, Appli
4	5	5.4	1279	3 US-09-248-335-25	Sequence 25, Appli
5	4	5.4	1416	3 US-08-866-928B-2	Sequence 2, Appli
6	7	5.4	1568	4 US-09-043-937A-1	Sequence 10, Appli
7	8	5.4	2167	4 US-09-489-869-10	Sequence 18, Appli
8	8	5.4	2249	3 US-08-860-635A-18	Sequence 18, Appli
9	8	5.4	2249	4 US-09-281-476-18	Sequence 18, Appli
10	8	5.4	3923	4 US-08-860-635A-20	Sequence 20, Appli
11	8	5.4	3923	4 US-09-281-476-20	Sequence 20, Appli
12	8	5.4	4060	1 US-08-308-949A-1	Sequence 1, Appli

13	8	5.4	5357	4 US-09-392-184-5	Sequence 5, Appli
14	8	5.4	4403765	4 US-09-103-840A-2	Sequence 2, Appli
15	8	5.4	4403765	4 US-09-103-840A-2	Sequence 2, Appli
16	8	5.4	4411529	4 US-09-103-840A-1	Sequence 1, Appli
17	8	5.4	4411529	4 US-09-103-840A-1	Sequence 1, Appli
18	7	4.7	24	1 US-08-411-796-187	Sequence 187, App
19	7	4.7	24	3 US-08-471-039-187	Sequence 187, App
20	7	4.7	24	3 PCT-US93-11198-187	Sequence 187, App
21	7	4.7	25	6 5504194-3	Patent No. 5504194
22	7	4.7	30	1 US-08-361-920-76	Sequence 76, Appli
23	7	4.7	30	1 US-08-479-939-76	Sequence 76, Appli
24	7	4.7	30	1 US-08-483-432-76	Sequence 76, Appli
25	7	4.7	30	5 PCT-US95-07372-3	Sequence 3, Appli
26	7	4.7	42	4 US-09-162-484-7	Sequence 3, Appli
27	7	4.7	51	2 US-08-350-260A-565	Sequence 565, App
28	7	4.7	53	1 US-08-081-539-59	Sequence 59, Appli
29	7	4.7	53	1 US-08-466-647-59	Sequence 59, Appli
30	7	4.7	54	1 US-07-941-651-3	Sequence 3, Appli
31	7	4.7	54	1 US-08-279-596-3	Sequence 3, Appli
32	7	4.7	54	1 US-08-814-220-26	Sequence 26, Appli
33	7	4.7	57	2 US-07-812-421-26	Sequence 26, Appli
34	7	4.7	57	4 US-09-043-303-14	Sequence 14, Appli
35	7	4.7	61	1 US-08-081-539-58	Sequence 58, Appli
36	7	4.7	61	1 US-08-466-647-58	Sequence 58, Appli
37	7	4.7	62	2 US-07-814-220-25	Sequence 25, Appli
38	7	4.7	62	2 US-07-812-421-25	Sequence 25, Appli
39	7	4.7	64	2 US-07-814-220-22	Sequence 22, Appli
40	7	4.7	64	2 US-07-812-421-22	Sequence 22, Appli
41	7	4.7	65	1 US-08-411-796-84	Sequence 84, Appli
42	7	4.7	65	1 US-08-411-796-85	Sequence 85, Appli
43	7	4.7	65	1 US-08-411-796-86	Sequence 86, Appli
44	7	4.7	65	1 US-08-411-796-87	Sequence 87, Appli
45	7	4.7	65	1 US-08-411-796-88	Sequence 88, Appli

ALIGNMENTS

RESULT 1
US-09-145-391-1
Sequence 1, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulist, Stefan M.
APPLICANT: Shubata, Hiroki
TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145.391
CURRENT FILING DATE: 1998-09-01
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (987)..
US-09-145-391-1

Alignment Scores:

Pred. No.: 2,39e-42 Length: 2372
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.51% Indels: 0
Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-145-391-1 (1-2372)

OY 49 PROSERGIEUVALTYTHSERALAMECPROGTYRPEPROTYRPPALALATHRALA 68
DB 1716 CCAGTTCACCTGTATATATATCTTGCAATGCAGGCTTCCTATCCAGCACGCCGC 1775

Qy 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyThrValGlyTrpThrPhe 88
Db 1776 GCGGCGCGCTACGAGGGGGGACCTGCGAGCGCGGTGCGACCTGACCACTTC 1835

Qy 89 ArgGlyAlaHisLeuArgGlyArgGlyThrValGlyValValTyr 105
Db 1836 AGGGCGCGCGCGCGCGCGCGCGATCCGGCTACGGCGGTGTGTTC 1886

RESULT 2

US-09-176-657-5
; Sequence 5, Application US/09176657
; Patent No. 6020164
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lu, Aina
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
; FILE REFERENCE: PF-0611 US
; CURRENT APPLICATION NUMBER: US/09/176,657
; EARLIER FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1250374
US-09-176-657-5

Alignment Scores:
Pred. No.: 0.00481 Length: 1506
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.78% Indels: 0
DB: 3 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-176-657-5 (1-1506)

Qy 73 ArgGlyAlaHisLeuArgGlyArgGlyThrValTyr 85
Db 999 AGAGAGAGCCCATTTGAGGGGAGAGGCGGACGATATAT 1037

RESULT 3

US-09-248-335-31
; Sequence 31, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEF, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 31
; LENGTH: 1007
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-31

Alignment Scores:
Pred. No.: 103 Length: 1007
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-248-335-31 (1-1007)

Qy 73 ArgGlyAlaHisLeuArgGlyArg 80
Db 234 CGAGGCGCTCACCTTCGAGACG 257

RESULT 4

US-09-248-335-25
; Sequence 25, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEF, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 25
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-25

Alignment Scores:
Pred. No.: 128 Length: 1279
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-248-335-25 (1-1279)

Qy 73 ArgGlyAlaHisLeuArgGlyArg 80
Db 196 CGAGGCGCTCACCTTCGAGACGA 219

RESULT 5

US-08-866-928B-2
; Sequence 2, Application US/08866928B
; Patent No. 6159702
; GENERAL INFORMATION:
; APPLICANT: Traish, Abdulmaged M
; TITLE OF INVENTION: In-vivo diagnostic methods for determining
; TITLE OF INVENTION: whether a primary breast tumor is clinically metastatic
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, P.C.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: MA
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB storage
; COMPUTER: IBM PS1
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,928B
; FILING DATE: May 30, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 509,570
; FILING DATE: July 31, 1995

ATTORNEY/AGENT INFORMATION:
NAME: David Prashker
REGISTRATION NUMBER: 29,693
REFERENCE/DOCKET NUMBER: BUMC-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (978) 525-3794
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-866-928B-2

Alignment Scores:
Pred. No.: 139 Length: 1416
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-08-866-928B-2 (1-1416)

QY 64 ProalaalathraAlaAlaAla 71
Db 87 CCAGCAGCAGCAGCAGCAGCAGC 110

RESULT 6
US-09-043-937A-1
Sequence 1, Application US/09043937A
Patent No. 6211432
GENERAL INFORMATION:
APPLICANT: BOUDET, ALAIN-MICHEL
PICHON, MAGALIE
GRIMA-PETTERATI, JACQUELINE
BECKERT, MICHEL
GAMAS, PASCAL
BRIAT, JEAN-FRANCOIS
TITLE OF INVENTION: DNA SEQUENCES CODING FOR CINNAMOYL-COA
REDUCTASE, AND APPLICATIONS THEREOF IN THE CONTROL OF
LIGNIN CONTENTS IN PLANTS
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE, P.C.
STREET: 110 NORTH GLEBE ROAD, 8TH FLOOR
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/043,937A
FILING DATE: 24-Jul-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR96/01544
FILING DATE: 03-OCT-1996
APPLICATION NUMBER: FR 95.11623
FILING DATE: 03-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 1568 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 278..1306
SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-043-937A-1

Alignment Scores:
Pred. No.: 153 Length: 1568
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-043-937A-1 (1-1568)

QY 64 ProalaalathraAlaAlaAla 71
Db 281 CCGCGCGCTACCGCAGCGCGCGCC 304

RESULT 7
US-09-489-869-10/c
Sequence 10, Application US/09489869A
Patent No. 6268151
GENERAL INFORMATION:
APPLICANT: Susan Murray
APPLICANT: Lex M. Cowser
TITLE OF INVENTION: ANTISENSE MODULATION OF MACROPHAGE MIGRATION INHIBITORY FACTOR
FILE REFERENCE: RTS-0110
CURRENT APPLICATION NUMBER: US/09/489,869A
CURRENT FILING DATE: 2000-01-20
NUMBER OF SEQ ID NOS: 88
SEQ ID NO 10
LENGTH: 2167
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1173)...(1280)
FEATURE:
NAME/KEY: CDS
LOCATION: (1470)...(1642)
FEATURE:
NAME/KEY: CDS
LOCATION: (1738)...(1804)
US-09-489-869-10

Alignment Scores:
Pred. No.: 203 Length: 2167
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-489-869-10 (1-2167)

QY 89 ArgAlaAlaAlaProProProPro 96
Db 1684 CGCGCGCGCGCGCGCTCCCGCGCC 1661

RESULT 8
US-08-860-635A-18/c
Sequence 18, Application US/08860635A
Patent No. 6143878
GENERAL INFORMATION:
APPLICANT: Koopman, Peter

APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,635A
FILING DATE: 29-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM9714
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2249 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-860-635A-18
Alignment Scores:
Pred. No.: 210 Length: 2249
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
Gaps: 0
DB: 3
US-09-809-545a-2_copy_1_148 (1-148) x US-08-860-635A-18 (1-2249)
Qy 89 ArgAlaAlaProProProPro 96
Db 788 CCGCAGCGCGCTCCGCTCTCCA 765
RESULT 9
US-09-281-476-18/c
Sequence 18, Application US/09281476
GENERAL INFORMATION:
APPLICANT: Koopman, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY

COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/281,476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/860,635
FILING DATE:
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2249 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-281-476-18
Alignment Scores:
Pred. No.: 210 Length: 2249
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
Gaps: 0
DB: 4
US-09-809-545a-2_copy_1_148 (1-148) x US-09-281-476-18 (1-2249)
Qy 89 ArgAlaAlaProProProPro 96
Db 788 CCGCAGCGCGCTCCGCTCTCCA 765
RESULT 10
US-08-860-635A-20/c
Sequence 20, Application US/08860635A
GENERAL INFORMATION:
APPLICANT: Koopman, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,635A

FILING DATE: 29-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM9714
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3923 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-860-635A-20

Alignment Scores:
Pred. No.: 342 Length: 3923
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-08-860-635A-20 (1-3923)

QY 89 ArgAlaAlaIaProProProPro 96
Db 846 CGCGAGCGGCTCGCGCTCTCCA 823

RESULT 11
US-09-281-476-20/c
Sequence 20, Application US/09281476
Patent No. 6316597
GENERAL INFORMATION:
APPLICANT: Koopman, Peter
APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/281,476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/860,635
FILING DATE:
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995

ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3923 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-281-476-20

Alignment Scores:
Pred. No.: 342 Length: 3923
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-281-476-20 (1-3923)

QY 89 ArgAlaAlaIaProProProPro 96
Db 846 CGCGAGCGGCTCGCGCTCTCCA 823

RESULT 12
US-08-308-949A-1/c
Sequence 1, Application US/08308949A
Patent No. 5580703
GENERAL INFORMATION:
APPLICANT: Kotin, Robert M.
APPLICANT: Berns, Kenneth I.
APPLICANT: Linden, Ralph M.
TITLE OF INVENTION: Human Adeno-Associated Virus Integration
TITLE OF INVENTION: Site DNA and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,949A
FILING DATE: September 20, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,127
FILING DATE: September 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC92-10F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-6240
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-08-308-949A-1

Alignment Scores:
Pred. No.: 353 Length: 4060
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 1 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-08-308-949A-1 (1-4060)
QY 89 ArgAlaAlaAlaProProProPro 96
DB 484 CCGGCAGCCGCCGCCGCCGCCGCCG 461

RESULT 13
US-09-392-184-5
/ Sequence 5, Application US/09392184
/ Patent No. 6395889
/ GENERAL INFORMATION:
/ APPLICANT: Robison, Keith E.
/ TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
/ FILE REFERENCE: 5800-55
/ CURRENT APPLICATION NUMBER: US/09/392,184
/ CURRENT FILING DATE: 1999-09-09
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: FASTSEQ for Windows Version 3.0
/ SEQ ID NO 5
/ LENGTH: 5357
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(5357)
/ OTHER INFORMATION: repolysin (NDAM family of metalloprotease)
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(5357)
/ OTHER INFORMATION: n = A, T, C or G
US-09-392-184-5

Alignment Scores:
Pred. No.: 450 Length: 5357
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-392-184-5 (1-5357)
QY 64 ProAlaAlaThAlaAlaAlaAla 71
DB 4638 CCGCGCGCACGCTGCGCGCTGCT 4661

RESULT 14
US-09-103-840A-2
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1,52e+05 Length: 4403765
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-103-840A-2 (1-4403765)
QY 64 ProAlaAlaThAlaAlaAlaAla 71
DB 2932874 CCGGCTGCACGCGCGCGCGCGCC 2932851

SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1,52e+05 Length: 4403765
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-103-840A-2 (1-4403765)
QY 90 AlaAlaAlaProProProProle 97
DB 2295213 GCCGCTGCACCGCCTCCACCGATC 2295236

RESULT 15
US-09-103-840A-2/C
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1,52e+05 Length: 4403765
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-103-840A-2 (1-4403765)
QY 64 ProAlaAlaThAlaAlaAlaAla 71
DB 2932874 CCGGCTGCACGCGCGCGCGCGCC 2932851

Search completed: March 16, 2003, 05:21:14
Job time : 1223 secs
```



```
Oy 1 MetThrAsnLysLeuValAlaIleProTyrThrAsnGlyTyrPheLeuIleProValVal 20
Db 535 ATGACTTAATAAAAGGCGCTGAACCCCTACACCAATGGCTGGAATTAATCCAGTTGTG 594
Oy 21 G1YA1AValTyrSerProIleProPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 595 GGGCGCGCTTACAGCCGCCACCTTCTATGACAGGACGGTCTGTGTGGCCAGGCGCAACAG 654
Oy 41 G1UG1SerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 655 GAGGATCTTCCATGATGACAGTGGCCCACTTCTATATCTTCTGCAATGCTTGGC 714
Oy 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 715 TTTCATATCCGGCGCCGCACTGCTGACGCTACGAGGAGGCGCTCACCTTCCAGGCCCT 774
Oy 81 G1YA1GThValTyrThrAsnThrPheAlaGlnAlaAlaAlaProProProProIleProAlaTyr 100
Db 775 GGTCCGACCGTGTACAAACCTTCAAGCTGCGCGCCGCCCAATCCCGGCTTAT 834
Oy 101 G1YGLYValValTyrGlnGlnProValTyrGlyAsnLysLeuGlnGlyGlyTyrAla 120
Db 835 GCGGAGTGTGTATCAAGACAGTATGCGCAATTAATGCTTACAGGCTGTTACGCT 894
Oy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
Db 895 GCATACCGCTACGCCCGACCCCTGCGACCTGCTGCTACAGTACAGTACAGTACGGA 954
Oy 141 ArgValTyrAlaAlaAlaAspProTyr 148
Db 955 CGAGTTTATGCTGCGACCCCTTAC 978

RESULT 2
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulec, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987)..(1979)
US-09-794-591-1

Alignment Scores:
Pred. No.: 8,48e-46 Length: 2372
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.51% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-794-591-1 (1-2372)
Oy 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
Db 1716 CCCAGTTCACTTGTATATCTTCTGCAATGCCAGGCTTCCCGATCCAGCAGCACCGCC 1775
Oy 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyValArgThrValTyrAsnThrPhe 88
```

```
Db 1776 GCGGCGGCTTACCGAGGCGGACCTGCGAGGCGCGGCTGCAACGCTGTACAAACCTTC 1835
Oy 89 ArgAlaAlaAlaProProProProIleProAlaTyrGlyGlyValValTyr 105
Db 1836 AGGCGCGCGCGCGCGCGCGCGCGCGCGATCCCGGCTACGGGGGTGTGTTAC 1886

RESULT 3
US-09-878-574-3090
; Sequence 3090, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3090
; LENGTH: 234
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-016-Q1-B1-B8
US-09-878-574-3090

Alignment Scores:
Pred. No.: 14.6 Length: 234
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-878-574-3090 (1-234)
Oy 64 ProAlaAlaThrAlaAlaAlaAla 71
Db 22 CCGGCGGCGCAGCGCGGCGGACGC 45

RESULT 4
US-09-923-876-1816/c
; Sequence 1816, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 1816
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700159573H1
; NAME/KEY: unsure
; LOCATION: 26, 34, 64, 93-94, 151
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-1816
```


Alignment Scores:

Pred. No.:	15.6	Length:	251
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.41%	Indels:	0
DB:	10	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-923-876-1816 (1-251)

QY 89 ArgAlaAlaAlaProProProPro 96

Db 122 CGAGCGCGCGCGCTCCGCGCT 99

RESULT 5

US-09-923-876-4839/C
; Sequence 4839, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Kamigaki, Ragnunath V.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/2298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 4839
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700455882H1
; LOCATION: 48, 236, 243, 245
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-4839

Alignment Scores:

Pred. No.:	16.3	Length:	263
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.41%	Indels:	0
DB:	10	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-923-876-4839 (1-263)

QY 64 ProAlaAlaThrAlaAlaAla 71

Db 104 CCTGCGCACTGCGGCTGCAGCG 81

RESULT 6

US-09-983-965-5536/C
; Sequence 5536, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Weeley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathiasen, Naeggeppan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26

; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5536
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 49-LIB34-024-Q1-E1-E2
US-09-983-965-5536

QY 64 ProAlaAlaThrAlaAlaAla 71

Db 132 CCAGCAGCAACAGCAGCGCGCC 109

RESULT 7

US-09-815-343-59
; Sequence 59, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-59

Alignment Scores:

Pred. No.:	22.2	Length:	368
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.41%	Indels:	0
DB:	10	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-815-343-59 (1-368)

QY 64 ProAlaAlaThrAlaAlaAla 71

Db 39 CCAGCAGCAACAGCAGCGCGCC 62

RESULT 8

US-09-864-761-32302
; Sequence 32302, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: GENE EXPRESSION ANALYSIS BY MICROARRAY

```
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32302
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC024123.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
OTHER INFORMATION: EST HUMAN HIT: BE872908.1, EVALUE 0.00e+00
OTHER INFORMATION: NT HIT: G11436757, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q07954, EVALUE 1.00e-23
US-09-864-761-32302

Alignment Scores:
Pred. No.: 23.4 Length: 389
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
Gaps: 0
DB: 10

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-864-761-32302 (1-389)
QY 64 ProAlaAlaThAlaAlaAlaAla 71
Db 324 CCGGGCCGCCACCGCTGCGGCTCG 347

RESULT 9
US-09-960-352-8373/C
Sequence 8373, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
```

```
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Machiagan, Naeppan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 8373
LENGTH: 407
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (359)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: 36-LIB3057-013-Q1-K1-A12
US-09-960-352-8373

Alignment Scores:
Pred. No.: 24.4 Length: 407
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
Gaps: 0
DB: 10

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-960-352-8373 (1-407)
QY 64 ProAlaAlaThAlaAlaAlaAla 71
Db 48 CCGGGCCGCCACCGCTGCGGCT 25

RESULT 10
US-10-025-380-659
Sequence 659, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasar A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 659
LENGTH: 411
TYPE: DNA
ORGANISM: Homo sapiens
US-10-025-380-659

Alignment Scores:
Pred. No.: 24.6 Length: 411
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-10-025-380-659 (1-411)

OY 64 Proba1a1aThra1a1a1a1a1a 71
DB 199 CCAGCAGCAACAGCAGCAGCGCC 222

RESULT 11

US-09-922-217-659
; Sequence 659, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugui
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 659
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-217-659

Alignment Scores:
Pred. No.: 24.6 Length: 411
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-922-217-659 (1-411)

OY 64 Proba1a1aThra1a1a1a1a1a 71
DB 199 CCAGCAGCAACAGCAGCAGCGCC 222

RESULT 12

US-09-833-263-659
; Sequence 659, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 659
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens

ORGANISM: Homo sapien
US-09-833-263-659

Alignment Scores:

Pred. No.: 24.6 Length: 411
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-833-263-659 (1-411)

OY 64 Proba1a1aThra1a1a1a1a1a 71
DB 199 CCAGCAGCAACAGCAGCAGCGCC 222

RESULT 13

US-09-864-761-11104/C
; Sequence 11104, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864.761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11104
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens

```

;
; FEATURE:
;
; OTHER INFORMATION: MAP TO AC002519.1
;
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
;
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.3
;
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
;
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
;
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
;
US-09-864-761-11104

Alignment Scores:
Pred. No.: 27.5 Length: 464
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-864-761-11104 (1-464)

OY 64 PROALAAATHALALALALAA 71
DB 228 CCTGCAGCAGCAGCAGCAGCA 205

RESULT 14
US-09-833-790-175
; Sequence 175, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongcong
; APPLICANT: Secrist, Heather
; APPLICANT: Monmarch, Raedoh
; APPLICANT: Indriats, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 175
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(501)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-790-175

Alignment Scores:
Pred. No.: 29.5 Length: 501
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-833-790-175 (1-501)

OY 90 AIAAIAAIAAIAAIAAIAAIA 97
DB 204 GCTGCTGCTGCTGCTGCTGCTCA 227

RESULT 15
US-09-864-761-15794
; Sequence 15794, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15794
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC024123.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
US-09-864-761-15794

Alignment Scores:
Pred. No.: 34.2 Length: 587
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-864-761-15794 (1-587)

OY 64 PROALAAATHALALALALAA 71
DB 393 CCGGCGGCGCGCGCTGCGGCGG 416

Search completed: March 16, 2003, 05:03:59
Job time : 100 secs
```


Oy 157 GATATTACTTGCAGCATGGCTGCTTCATTTCCGGGCC----- 198

Dbe 231 VALAAlaAgSerAlaMetHsrProTyrrProTyrrProCysSerArgLeuAlaIglYAsrSer 250

Oy 199 -----ACTGTGCAGCTGCATACCGAGGGGCTCACTTGAGAGCCGGTGC--- 246

Dbe 251 AlaAtgArGrPrCoAlAlaGlYArGrArGrAlYAlaArGrLeuAlaGrThnrlEglYTrCYs 270

Oy 247 ACCGTGTCAACAACC---TTCAGAGCTGCGGGGGCCCCAACCCCAATCCGGCCATAGGC 303

Dbe 271 AlaAlaTYgrlThleuYrAlaAlaAlaGlYglYglYAlaSerProSprtyrAsn 290

Oy 304 GGAGTAGTGTATCAAGACAGCCAGCTGTATGGCAATAATTGCTCACAGGGTGTATCCCTGCA 363

Dbe 291 SerSerGlUTyrGlUnglYProLeuSerProProLeuCyYLeuAlaMnglYAsnPhseThrleu 310

Oy 364 TACCCTACAGCCAGCCACCCCTGCCTGCCTGCTGCTGCCTACAGTGAACATTACGGAGA 423

Dbe 311 LysGIAsnAprSerSerProAprHrIaGlYAsSerTYrHlaTYrSerMeChlertYserAla 330

Oy 424 GTTATGCTGCCGACCCCTACACCAACACACTT 456

Dbe 331 LeuProGIYSerArGrPrCoAlaGlYHsrIleu 341

RESULT 4

A40020
collagen alpha 1(XII) chain precursor - chicken

N:Alternate names: fibrochimerin
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Jan-2001
C:Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
R:Yamaguchi, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obata
J. Cell Biol. 115, 209-221, 1991

A:Title: The complete primary structure of type XII collagen shows a chimeric molecule v
nous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site
A:Reference number: A40020; MUID:92011862; PMID:1918137

A:Accession: A40020
A:Molecule type: mRNA
A:Residues: 1-3124 <YAM>

A:Cross-references: GB:D00824; NID:g222810; PIDN:BA400701.1; PID:g222811
A>Note: In the author translation residues 1216-1219 are shown after residue 1235 and
R.Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989

A:Title: Type XII collagen. A large multidomain molecule with partial homology to type I
A:Reference number: A34485; MUID:90062079; PMID:2584192

A:Accession: A34485
A:Molecule type: mRNA
A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
A:Cross-references: EMBL:J05137; NID:g211284; PIDN:AAA48635.1; PID:g211285
A:Accession: B34485

A:Molecule type: protein
A:Residues: 2772-2792;2846-2873 <GOR2>
R.Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987

A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c
A:Reference number: A28037; MUID:87317590; PMID:3476925

A:Accession: A28037
A:Molecule type: mRNA
A:Residues: 2360-2376, 'F', 2978-3074, 'AG' <GOR3>
A:Cross-references: EMBL:M17375; NID:g211649; PIDN:AAA48718.1; PID:g211650
A>Note: This sequence has been revised in reference A34485
R.Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992

A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A:Reference number: S23814; MUID:92362621; PMID:1323460

A:Accession: S23814
A:Molecule type: protein
A:Residues: 'X',1333,'Q',1335-1347;1914-1928;2504,'X',2506,'X',2508-2511,'X',2513-2517 <
J. Biol. Chem. 262, 17724-17727, 1987

A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin
A:Reference number:S22254; MUID:88087065; PMID:3121603

A:Accession: 522254
A:Molecule type: Protein
A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843;3002-3014 <DUB>
R:Trnab, U.: Trnab, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A:Title: The two splice variants of collagen XII share a common 5' end.
A:Reference number: S28811, MUID:93042014, PMID:1420368
A:Accession: 52811
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24,1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>
A:Cross-references: EMBL:X67327
C:Genetics:
A:Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2932/1; 2985/1; 3008/1; 3065/1
C:Superfamily: collagen alpha 1(XII) chain, fibronectin type III repeat homology; von Willebrand factor type A repeat homology; cell binding; coiled coil, connective tissue, disulfide
C:Keywords: alternative splicing; cell binding; coiled coil, connective tissue, disulfide
F:1-23/Domains: signal sequence #status predicted <SIG>
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
F:24,1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted <IITA>
F:24-114/Domains: IIRa #status predicted <IIRa>
F:24-105/Domains: fibronectin type III repeat homology <FN3A>
F:1137-301/Domains: von Willebrand factor type A repeat homology <VWA1>
F:332-425/Domains: IIRB #status predicted <IIRB>
F:333-414/Domains: fibronectin type III repeat homology <FN3B>
F:437-601/Domains: von Willebrand factor type A repeat homology <VWA2>
F:629-1178/Domains: IIRIC #status predicted <IIRIC>
F:630-711/Domains: fibronectin type III repeat homology <FN3C>
F:721-802/Domains: fibronectin type III repeat homology <FN3D>
F:812-895/Domains: fibronectin type III repeat homology <FN3E>
F:905-986/Domains: fibronectin type III repeat homology <FN3F>
F:995-1076/Domains: fibronectin type III repeat homology <FN3G>
F:1086-1169/Domains: fibronectin type III repeat homology <FN3H>
F:1197-1361/Domains: von Willebrand factor type A repeat homology <VWA3>
F:1384-1295/Domains: IIRID #status predicted <IIRID>
F:1384-1465/Domains: fibronectin type III repeat homology <FN3I>
F:1474-1557/Domains: fibronectin type III repeat homology <FN3J>
F:1566-1647/Domains: fibronectin type III repeat homology <FN3K>
F:1655-1738/Domains: fibronectin type III repeat homology <FN3L>
F:1756-1838/Domains: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domains: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domains: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domains: fibronectin type III repeat homology <FN3P>
F:2119-2199/Domains: fibronectin type III repeat homology <FN3Q>
F:2207-2294/Domains: fibronectin type III repeat homology <FN3R>
F:2325-2490/Domains: von Willebrand factor type A repeat homology <VWA4>
F:2438-2440/Domains: cell adhesion #status predicted
F:2509-2750/Domains: IXP, homologous to NC4 domain of type IX collagen #status predicted <COL2>
F:2751-2902/Domains: collagenous COL2 #status predicted <COL2>
F:2899-2901/Region: cell attachment (R-G-D) motif
F:2903-2945/Domains: non-collagenous NC2 #status predicted <NC2>
F:2946-3048/Domains: collagenous COL1 #status predicted <COL1>
F:3049-3124/Domains: non-collagenous NC1 #status predicted <NC1>
F:32, 1006, 1032, 1044, 1512, 1767, 2210, 2273, 2532, 2683/Binding site: carboxydrate (Asn) (cov)
F:2780, 2789, 2836, 2842, 2860, 2866, 3004, 3007/Modified site: hydroxyproline (Pro) #status

Db 154 ThrIleuArgProTyrProSerSerValIleuIleCysSerIleuThrThrProSerThr 173
Qy 194 CGGCACGCTGCTGACGTGATACCGAGGGGCTCACCCTTGAGCGCGTGTGCACCG--- 250
Db 174 AlaProGlyGluGluValIleArgGlyIleArgTyrSerSerSerIleProSer 193
Qy 251 -----TGACACACCTTCAGAGCTGCGG 274
Db 194 SerAlaProProGluThrAlaAspAspProProGluCysLeuThrProThr----- 210
Qy 275 CGCCCCACCCCAATCCCGGCTATGGCGAGTAGTATCAAGACCGAGTATGCA 334
Db 211 ---ProThrProLeuSerThrProSer-----IleIleArgProGluProGluAla 226
Qy 335 ATAAAT-----TGCTACAGGGTGTT-----ACG 358
Db 227 LeuThrLeuProLeuGlySerAlaValSerGluThrThrProThrProIleSerThr 246
Qy 359 CTGCATACCGCTACCGCCAGCCCA---CCCTGCACCTGCTGCTACGCTACGACGA--- 412
Db 247 ThrThrThrValThrProThrProThrProThrGlyThrGluThrProThrThrPro 266
Qy 413 GTTACGACGAGTTATAGCTGCGACCCCTTACCAACACACTGTGCTCCAGCCCACT 472
Db 267 ILeuThrThrThrValThrProThrProThrProThrGly---ThrGluThrProThr 285
Qy 473 AGCGCTTGCTGCATGATGCTTTGGCCCTTCGACCGATGCCA 517
Db 286 ThrValLeuIleThrThrThrThrThrMetProThrProThrPro 300

RESULT 2

C75607 conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: C75607
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.D.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaluski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:2003696; PMID:10567266
A:Accession: C75607
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIRN:AA12236.1; PID:g646053
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0120
A:Map position: 2
C:Superfamily: ymbw protein

Alignment Scores:
Pred. No.: 0.649 Length: 338
Score: 103.00 Matches: 45
Percent Similarity: 37.02% Conservative: 22
Best Local Similarity: 24.86% Mismatches: 78
Query Match: 8.85% Indels: 36
DB: Gaps: 8

US-09-809-545A-1_copy_535_1143 (1-609) x C75607 (1-338)

Qy 58 GTGGGCGGCTTACAGCCCGACTTATGACAGCAGCGTCTGTTGTCACAGCCCAAC 117
Db 105 LeuGlyLeuGlyValArgAlaProGlyThrAspGlyArgThrAlaLeuAlaLeuArgSer 124
Qy 118 CAGGAGGATCTTCATGATACAGTGGCCAGTTCATATATATCTTCAATAGCCCT 177
Db 125 ArgGluAlaLeuAlaAlaAspAspPheProThrGluIleAlaLeuLeuAlaGluAla 144
Qy 178 GGC---TTTCATATCCGGCCGACCTGTGACGTGACATACCGAGGGCTCACCTTGA 234

Db 145 GlyIuThrProTyrPro-----AlaGlySerValPheAspGlyValLeuAlaGlu 161
Qy 235 GGGCGTGTGCGACCGGTATACACACCTTACAGCTGGCGGCCCCCAATCCCG 294
Db 162 ProArgGlyValIleValLeu-----ProIleuThr 171
Qy 295 GCTATGCGGAGTAGTATATCAAGACCGAGTATAGCAATTAATGCTACAGGGTGT 354
Db 172 IleuGlySerSerLeuThrGlyAlaGluLeuAlaGlyGlu-----LeuGlyValGly 189
Qy 355 TACGCT---GCATACCGTATACGCCAGCCCACTGCTGCTGCTGCTACAGTAC 411
Db 190 TyrAlaPheAlaIleValIlePheSerGluIleAspProAlaVal----- 203
Qy 412 AGTTACGACGAGTTATGCTGCGGACCCCTACACACACACTGTGTCAGCCCC--- 468
Db 204 -----AlaValGluThrTyrIleAspGlyPheArgProGlyProLeu 217
Qy 469 -----ACCTACGCGCTGTGTCATGATAGTCTTTGCGCCCTTGACCGATGCCAAGCT 522
Db 218 AlaGluProTyrAlaIleLeuGlyValAlaAlaLeuAla-----AlaProThr 233
Qy 523 AGGACCATGCTGATGATGAGGCTGCTCTTCTTCTTATTCAGAGCTAGTATACCA 582
Db 234 GluAspGluAlaArgAspLeuSerIleuThrSerAlaAlaLeuSerLeuGlyIleLeuSer 253
Qy 583 GGG 585
Db 254 Gly 254

RESULT 3

JC4647 K18 protein - rat

N:Alternate names: basic helix-loop-helix protein homolog
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C:Accession: JC4647
R:Kume, H.; Matuyama, K.; Tomita, T.; Iwatsubo, T.; Saido, T.C.; Obata, K.
Biochem. Biophys. Res. Commun. 219, 526-530, 1996
A:Title: Molecular cloning of a novel basic helix-loop-helix protein from the rat brain.
A:Reference number: JC4647; MUID:96193685; PMID:8605021
A:Accession: JC4647
A:Molecule type: DNA
A:Residues: 1-381 <KUM>
A:Cross-references: DDBJ:D82868; NID:g1163397; PID:BA11615.1; PID:g1163398
A:Experimental source: brain
C:Comment: This protein is involved in synaptic plasticity, and has a role specific to n
-loop-helix domain.
C:Keywords: brain
F:122-134/Region: basic

Alignment Scores:
Pred. No.: 0.645 Length: 381
Score: 103.00 Matches: 39
Percent Similarity: 36.84% Conservative: 24
Best Local Similarity: 22.81% Mismatches: 76
Query Match: 8.85% Indels: 32
DB: Gaps: 4

US-09-809-545A-1_copy_535_1143 (1-609) x JC4647 (1-381)

Qy 40 TGAATTAATCAAGTGTGTGGCGCGCTACACCCCGACTTCTTACAGCAGCGTG 99
Db 171 TrpAlaLeuSerGluIleuArgSerGlyValArgProAspLeuValSerTyrValGlu 190
Qy 100 CTGTGTGTCAGGCC----- 114
Db 191 ThrLeuCysGlyLeuSerGluProThrThrAsnLeuValAlaGlyCysLeuGluLeu 210
Qy 115 -----AACGAGAGGATCTTCATGATACAGTGGCCCGCCAGTTCACCT 156
Db 211 AsnSerArgAsnPheLeuThrThrGluGluAlaAspGlyValAlaAlaPheThrAlaArg 230

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 12, 2003, 19:22:28 ; Search time 23 Seconds
(without alignments)
5090.945 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143
Sequence: 1 atgactataataaaagccgct.....acaacgcttcgctccatcat 609

Scoring table:
BLOSUM62 Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2.1/USFPO.spool/US09809545/runac_11032003_130946_12825/app.query.fasta_1.775
-DB=PIR_73 -QWMT=fastcan -SUFFIX=n2p.rpr -MINMATCH=0.1 -T00PCT=0 L00PEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09809545 @CNC 1.1 25 @runac_11032003_130946_12825 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-MARN TIMEOUT=30 -THEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: PIR_73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	9.3	1118	2	A48292
2	103	8.8	338	2	C75607
3	103	8.8	381	2	JC4647
4	103	8.8	3124	1	A40020
5	102.5	8.8	241	2	T22216
6	101	8.7	1221	2	T73472
7	100.5	8.6	211	2	B89716
8	99	8.5	387	2	I50703
9	99	8.5	542	2	T29707
10	98.5	8.5	655	1	A55726
11	98.5	8.5	2639	2	T13328
12	98	8.4	216	2	I51920
13	98	8.4	316	2	T31880
14	98	8.4	376	2	A26066

15	97	8.3	811	2	T36581	probable transmem
16	96.5	8.3	139	2	S61885	extensin precursor
17	96.5	8.3	748	2	T04011	hypothetical prote
18	96	8.2	405	4	A61181	homeotic protein H
19	96	8.2	760	2	A45174	eye cell developme
20	95.5	8.2	753	2	J00532	OP protein - Kenne
21	95.5	8.2	1172	1	TSHUP2	chromospondin 2 p
22	95	8.2	418	2	T15142	hypothetical prote
23	94.5	8.1	268	2	B61615	fibroin heavy chai
24	94.5	8.1	469	2	T20047	hypothetical prote
25	94.5	8.1	1006	2	G86292	hypothetical prote
26	94.5	8.1	1964	2	T09059	notch4 - mouse
27	94	8.1	219	2	S42574	adhesive protein -
28	94	8.1	967	2	H86334	T20H2.10 protein -
29	93.5	8.0	294	2	A37232	mucin, tracheal (A
30	93.5	8.0	320	2	S09208	chorion protein s3
31	93.5	8.0	339	2	T25562	hypothetical prote
32	93.5	8.0	839	2	T04859	extensin homolog F
33	93	8.0	197	2	T10081	sperm mitochondria
34	93	8.0	495	1	S31223	transcription fact
35	92.5	7.9	164	2	T15525	hypothetical prote
36	92.5	7.9	348	2	T35248	probable oxidoredu
37	92.5	7.9	656	1	A49358	RNA-binding protei
38	92	7.9	250	2	C87443	conserved hypochet
39	92	7.9	443	2	I49140	p62 ras-GAP associ
40	92	7.9	535	2	S65762	chitinase (EC 3.2.
41	91.5	7.9	325	2	C95952	hypothetical prote
42	91.5	7.9	409	2	A70932	probable PEP prote
43	91.5	7.9	856	2	T13159	EB1-55kDa-associat
44	91.5	7.9	1052	2	T14343	zinc finger RNA bi
45	91.5	7.9	2715	2	T13049	eyelid - fruit fly

ALIGNMENTS

RESULT 1
A48292
mucin, tracheobronchial - dog
C/Species: Canis lupus familiaris (dog)
C/Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 05-Nov-1999
C/Accession: A48292
R/Verma, M.; Davidson, E.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 7144-7148, 1993
A/Title: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA contair
A/Reference number: A48292; MUID:93348229; PMID:8346228
A/Accession: A48292
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1118 <VER>
A/Cross-references: GB:X69164; NID:9402557; PIDN:CAA48914.1; PID:9402558
A/Note: authors translated the codon CAA for residue 52 as Gly, GAT for residue 94 as Asr
C/Comment: It is uncertain whether Met-1 or Met-52 is the initiator.
C/Keywords: glycoprotein

Alignment Scores:		Pred. No.:	0.228	Length:	1118
Score:		Percent Similarity:	108.00	Matches:	57
Best Local Similarity:		Best Local Similarity:	40.61%	Conservative:	23
Query Match:		9.28%	Mismatches:	76	
DB:		2	Indels:	41	
			Gaps:	10	
US-09-809-545A-1_COPY_535_1143 (1-609) x A48292 (1-1118)					
QY	17	CCGCGAACCCCTACACCAATGCTGGAATTAATCCAGTTGGGCGGCTACAGCC	76		
DB	117	ProGlnThrProThrProLeuSer-ThrProAlaSerSerSerIlyPro-----GlnPr	134		
QY	77	CCGACTTCATGACGAGCGGTCTGTGTCACGACCAACGAGGAGATCTTCATGT	136		
DB	134	OHs-PheTyPProSerSerValIeuSerAlaValSerGluThrThProThrThr-Thr	153		
QY	137	ACAGTGCCGCCAGTTCATCTGTATATATCTTCGCAATGC---CTGGCTTTCATATCCGG	193		

THIS PAGE BLANK (USPTO)

```

FN W0200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
XX N-PSDB: AAK53229.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 423; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, hematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX
XX SQ Sequence 286 AA;
XX
XX Alignment Scores:
XX Pred. No.: 3,866-16 Length: 286
XX Score: 258.00 Matches: 47
XX Percent Similarity: 96.08% Conservative: 2
XX Best Local Similarity: 92.16% Mismatches: 2
XX Query Match: 22.16% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545A-1_COPY_535_1143 (1-609) x AAM80096 (1-286)
XX
XX 1 ATGACTAATATAAGGCGGCTGAAACCCCTACACCAATGGCTGGAATTAATCCAGTTGG 60
XX |||||||
XX Db 219 MetTrhAmLbVbLpThrValAsnProTyrTrhAsnGlyTrpLysLeuAsnProValVal 238
XX |||||||
XX 61 GGCGGGGTCTACAGCCCGACTTCTATGAGGACAGCGTGCTGTGTGCCAGGCCAACG 120
XX |||||||
XX Db 239 GlyAlaValIlySerProGluPheTyrAlaGlyThrValLeuLeuCyGlnIlaAsnGln 258
XX |||||||
XX 121 GAGGATCTTCATGATACAGTGCGCCCACTTCA 153
XX |||||||
XX Db 259 GluGlySerSerMetLysSerAlaProSerThr 269
XX |||||||

```

Search completed: March 12, 2003, 21:20:40
 Job time : 35 secs

QY 514 GCCAAGACTAGAGCCAGCTGATGATGGGTCTGTTCTTTCTTATTCAGAGCTAGT 573
Db 340 uPro---LeuAlaMetGluLeuAlaLeuTrip-----ArgVa 351
QY 574 ATATPCCAGGGGATACACCGTTTG 601
Db 351 ITyrtHrGluValAlaThrAlaAspLeu 360
RESULT 14
ABG26323
ID ABG26323 standard; Protein, 388 AA.
AC ABG26323;
XX 18-FEB-2002 (first entry)
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #26314.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS90510.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 20; SEQ ID NO 56682; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 388 AA;
SQ

Alignment Scores: 3.25e-21 Length: 388
Pred. No.:

Score: 311.50 Matches: 90
Percent Similarity: 46.64% Conservative: 14
Best Local Similarity: 40.36% Mismatches: 59
Query Match: 26.76% Gaps: 60
DB: 22 Indels: 10
US-09-809-545A-1_COPY_535_1143 (1-609) x ABG26323 (1-388)
QY 1 ATGACATATAAAGGCGGTGAAACCCCTACACCAATGGCTGGAATTAATTCAGTTGTG 60
Db 186 MetThrAnuLysMetValThrProTyrAlaAsnGlyTrpLysLeuSerProValVal 205
QY 61 GCGCGGCTTACAGCCCGGACTTATGACAGGACGCGGTGCTGTGCCAGGCCAAC--- 117
Db 206 GlyAlaValTyrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu 225
QY 118 ---CAGGAGGATCTTCATGATGACGTGCGCCCGGCTTCACTT-----GTATATACCTTCT 168
Db 226 GlyLeuAspAlaAlaValProLeuSerGlyArgGlyLeuAlaThrTyrIleProLeu 245
QY 169 GCAATGCGCTGCTTTCATATCCG---GCCGCCACTGCTGCAGCTGCATACCGAGGGCT 225
Db 246 IleIleProGlyPheProTyrProThrAlaAlaThrThrAlaAlaAlaPheArgGlyAla 265
QY 226 CACCTTGAGCGCGGTGTCGACCGGTGACACACCTTCAGAGCTGGCGCGGCCCAACC 285
Db 266 HisLeuArgGlyArgGlyArgThrValTyrGlyAlaValArg---AlaValProProThr 284
QY 286 CCAATCCCGGCTATGGCGGAGTA----- 309
Db 285 AlaIleProAlaTyrProGlyValAlaAspLeuProGlyThrAspPheThrValLeuThrSer 304
QY 310 ---GTGTATCAAGACCCAGTGTATGCAATTAATGTCTACAGGGGTGTTACGCTGCATAC 366
Db 305 MetValAspMetGlnProThrAspMetHisSerLeuLeu----- 317
QY 367 CGCTACGCCGACGCCACCCCTGCTGCTGCT-----ACAGTGACAGT 414
Db 318 ----LeuGlnProGlnPro-ProLeuLeuGlnProLeuGlnProLeuThrValThrVa 335
QY 415 TACGAGCAGATTATGCTGCCGACCCCTACACCCACACACTTGCTCCAGCCCCACCTAC 474
Db 335 lMetAlaGlyCysThrGlnProThr----- 345
QY 475 GCGCTTGCGCATGATGATCTTTGCGGCTTGACCGATGCCAGACTAGAGCCATGCT 534
Db 346 -----MetProLeuProLeuPro---LeuAlaMetGluLe 356
QY 535 GATGATGCGGCTGCTTCTTTCTTCTTATTCAGAGCTAGTATATACAGGGGATACAAAC 594
Db 356 uAlaLeuTrip-----ArgValTyrThrGluValAlaThrAl 368
QY 595 CGTTTG 601
Db 368 aspLeu 370
RESULT 15
AAM80096
ID AAM80096 standard; Protein, 286 AA.
AC AAM80096;
XX 06-NOV-2001 (first entry)
DT Human protein SEQ ID NO 3742.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
OS

Qy	412	TACGGCGTTGGTGCATGAATGCTTTTGGCGCCCTTGACCCATGCCAAGACTAGAGCCAT	531
Db	24	TyrgLyValGIgLYAlaMetAsnAlaPheAlaProLeuThrIspAlaLyStrIArgSerHis	43
Qy	532	GCTGATGATGTGGGCTCGTCTCTTTCTTCATTTGACAGGCTAGTATATACCAAGGGGATAC	591
Db	44	AlaAspAspValGIgLYLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGIgLYTyr	63
Qy	592	AACCGTTTGCTGCATAT	609
Db	64	AsnArgPheAlaProTyr	69
RESULT 13			
ID	AA156850	standard; Protein; 366 AA.	
XX	AA156850;		
AC	AA156850;		
XX	11-APR-2000	(first entry)	
DT	11-APR-2000	(first entry)	
XX	Human RNA binding protein (RNABP)-2 (clone 1250374).		
DE	RNA binding protein; RNABP; cancer; immune disorder; AIDS; human;		
KW	developmental disorder; acquired immunodeficiency syndrome; RNABP-2;		
KW	inflammation; allergy; diabetes mellitus; seizure disorder.		
XX			
XX	Homo sapiens.		
OS			
XX	US6020164-A.		
PN	US6020164-A.		
XX	01-FEB-2000.		
PD	01-FEB-2000.		
XX	21-OCT-1998;	98US-0176657.	
PF	21-OCT-1998;	98US-0176657.	
XX			
PR	21-OCT-1998;	98US-0176657.	
XX			
PA	(INCY-) INCYTE PHARM INC.		
PI	Bandman O, Corley NC, Lu DM, Baughn MR, Tang YT, Guegler KJ;		
XX	WPI; 2000-146885/13.		
DR	N-PSDB; AA246827.		
XX			
PT	Isolated and purified polynucleotide for modulating the expression of		
PT	human RNA binding proteins which play a role in cancer, immune		
PT	disorders and developmental disorders -		
PS	Examples; Columns 45-48; 39pp; English.		
XX			
CC	The present invention provides human RNA binding proteins (RNABP) and		
CC	polynucleotides encoding the proteins. The polypeptides are useful for		
CC	modulating the expression of human RNA binding proteins (RNABP) which		
CC	play a role in cancer, immune disorders and developmental disorders.		
CC	Disorders associated with a decrease of RNABP include: cancers such as		
CC	adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,		
CC	teratocarcinoma, and, in particular, cancers of the adrenal gland,		
CC	bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia		
CC	gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,		
CC	pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,		
CC	testis, thymus, thyroid, and uterus, immune disorders such as acquired		
CC	immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory		
CC	distress syndrome, allergies, ankylosing spondylitis, amyloidosis,		
CC	anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune		
CC	thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's		
CC	disease, atopic dermatitis, dermatomyositis, diabetes mellitus,		
CC	emphysema, episodic lymphopenia with lymphocytocoxins, erythroblastosis		
CC	falleis, erythema nodosum, atrophic gastritis, glomerulonephritis,		
CC	Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,		
CC	hyperosteoarthritis, irritable bowel syndrome, multiple sclerosis,		
CC	myasthenia gravis, myocardial or pericardial inflammation,		
CC	osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,		
CC	Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's		

CC	syndrom, systemic anaphylaxis, systemic lupus erythematosus, systemic
CC	scleriosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner
CC	syndrome, complications of cancer, hemodialysis, and extracorporeal
CC	circulation, viral, bacterial, fungal, parasitic, protozoal, and
CC	helminthic infections, trauma, x-linked agammaglobulinemia of Bruton,
CC	common variable immunodeficiency (CVI), digoxin's syndrome (thymic
CC	hypoplasia), thymic dysplasia, isolated immunoglobulin (Ig)-A deficiency,
CC	severe combined immunodeficiency disease (SCID), immunodeficiency with
CC	thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi
CC	syndrome, chronic granulomatous diseases, hereditary angioneurotic edema,
CC	and immunodeficiency associated with Cushing's disease; and developmental
CC	disorders such as renal tubular acidosis, anemia, Cushing's syndrome,
CC	achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,
CC	epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia,
CC	genitourinary abnormalities, and mental retardation), Smith-Magenis
CC	syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis,
CC	hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-
CC	Toutz disease and neurofibromatosis, hypothyroidism, hydrocephalus,
CC	seizure disorders such as Sydenham's chorea and cerebral palsy, spina
CC	bifida, anencephaly, craniofacial dysmorphism, congenital glaucoma, cataract,
CC	sensorineural hearing loss, and any disorder associated with cell
CC	growth and differentiation, embryogenesis, and morphogenesis involving
CC	any tissue, organ, or system of a subject, e.g., the brain, adrenal
CC	gland, kidney, skeletal or reproductive system. The present sequence
CC	represents the RMBP-2 polypeptide.
XX	
XX	Sequence 366 AA;
XX	
Alignment Scores:	
Pred. No.:	1,326-21 Length: 366
Score:	315-50 Matches: 90
Percent Similarity:	49.05% Conservative: 13
Best Local Similarity:	42.864 Mismatches: 59
Query Match:	27.10% Indels: 48
DB:	21 Gaps: 10
US-09-809-545A-1_COPY_535_1143 (1-609) x AAY56850 (1-366)	
Oy	1 ATGACTATATAAAGCGCGCTGAACCCCTACACCAATGCGTGAATTAATTCAGTTGTG 60
Db	
Db	190 MetrhmrhnylsyhsmeValThrProTyrAlaAenGlyTrpIlyLeuSerProValVal 209
Oy	61 GGGCGGCTTACAGCCCGACCTTATGCAAGCAGCGTGGCTGTGSCAGGCCAAC-- 117
Db	
Db	210 GlyAlaValTyrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu 229
Oy	118 ---CAGAGGAGATCTTCACATGACACTGGCCCCACGTCACTT-----GTATATACCTCT 168
Db	
Db	230 GlyAenAspAlaAlaValProLeuSerGlyArgGlyGlyIleAsnThrTyrIleProLeu 249
Oy	169 GCATGCGCTGCTTCCATATCCG---GGCGCACAGTGCAGCTGATACCGAGGGCT 225
Db	:::
Db	250 IleIleProGlyPheProTyrProThrAlaAlaThrThrAlaAlaPheArgGlyAla 269
Oy	226 CACCTTGAGAGCGGTGATGCGACCGCTGTACAACACCTTGAAGCTGGCGCCCCACCC 285
Db	
Db	270 HisLeuArgGlyArgGlyArgThrValTyrGlyAlaValArg--AlaValProProThr 288
Oy	286 CCAATCCCGGCTATGGCGGAGTGAAGTATCAAGACCCAGTATGCAATTAATTGTA 345
Db	
Db	289 AlaIleProAlaTyrProGlyValaAspMetGln---ProThrAspMetHisSerLeuLeu 307
Oy	346 CAGGTGTTAAGCTGATACCGCTACCGCCGACCGCCACCCCTGCAGCTGCTGCTCT-- 403
Db	
Db	308 -----LeuGlnProGlnPro-ProIlyLeuGlnProIle 318
Oy	404 -----ACAGTACAGTTACGAGCAGAGTTTATGCTGCGACCCCTACACACACA 453
Db	
Db	318 uGlnProLeuThrValThrValMetAlaGlySerThrGlnProThrProThr----- 335
Oy	454 CTTGCTCAGACCCCACTAAGCGCGTGTGCTGCATGAATGCTTTTGCGCCCTTACCGAT 513
Db	
Db	336 -----MetProIlyProIle 340

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6810; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA438642-AA442213) with neurotrophic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, Leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 291 AA:
SQ
Alignment Scores:
Pred. No.: 1.18e-28 Length: 291
Score: 389.00 Matches: 90
Percent Similarity: 57.93% Conservative: 5
Best Local Similarity: 54.88% Mismatches: 38
Query Match: 33.42% Indels: 31
DB: Gaps: 4
US-09-809-545A-1_COPY_535_1143 (1-609) x AA441879 (1-291)
QY 1 ATGACTAATAAAGCGCGTGAACCCCTACACCAATGCTGGAATTAATCCAGTTG 60
DB 147 MetThrAsnLysLysThrGlyAsnProTyrThrAsnGlyTyrLysLeuAsnProValVal 166
QY 61 GCGCGGCTGTACAGCCCGACCTTCTAAGCAGGCGGCTGCTTGGCAGGCCAACCG 120
DB 167 GlyAlaValValTyrGlyProGluPheTyrAlaValThr----- 178
QY 121 GAGGATCTTCATGATACAGTGCGCCAGTTCATTGTATATCTTCGAAATGCGTGC 180
DB 179 -----Gly 179
QY 181 TTTCATATCGGCGCGCAGCTGCTGAGCTGCATACCGAGGGGCTCAGCTTGAGGCGCT 240
DB 180 PheProTyrProThrThrGlyThrAlaValAlaTyrAlaGlyAlaHisLeuArgIleArg 199
QY 241 GGTCCGACCGCTGACAAACCTTCAGAGCTGGCGCGCCGCCCAATCCCGCTAT 300
DB 200 GlyAlaGlyAlaValTyrAsnThrPheArgAlaAlaProProProProIleProThrTyr 219
QY 301 GCGGAGTGTATGATCAAGAGCAGTGTGCAATAAATGCTACAGGGGTGTTACGCT 360
DB 220 GlyAlaValValTyrGlnAspGlyPheTyrGlyAlaGluIleuAla----ThrG 238
QY 361 GCATACCGGTACGCCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 238 nProThrArgThrLeuSer---ProLeuGlnArgGlnProThrAlaThrValThrAl 257
QY 421 CGAG---TTTATGCTGCCAGCCCTTCACCAACACACTTGTCTCCAGCCCGCTACGG 477
DB 257 agLuserrThnGlnLeuProThrArgThrIleThrProSerGlyProArgArgProThrAl 277
QY 478 GTTGATGCCA 487
DB 277 aleuGluPro 280
RESULT 12
ABG29445

ID ABG29445 standard; Protein; 103 AA.
XX
XX AC ABG29445;
XX
XX 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #29436.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS93632.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID NO 59804; 103bp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 103 AA:
Alignment Scores:
Pred. No.: 2.72e-24 Length: 103
Score: 342.00 Matches: 64
Percent Similarity: 98.48% Conservative: 1
Best Local Similarity: 96.97% Mismatches: 1
Query Match: 29.38% Indels: 0
DB: Gaps: 0
US-09-809-545A-1_COPY_535_1143 (1-609) x ABG29445 (1-103)
QY 412 AGTTACGCGAGGATTATGCTGCCAGCCCTTACCAACACACTTGTCTCAGCCCGCACC 471
DB 4 SerTyrGlyArgValTyrAlaAlaAspProTyrThrHisAlaLeuAlaProIleProThr 23

RESULT 9

ABG29447

ID ABG29447 standard; Protein: 705 AA.

AC ABG29447;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #29438.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX N-PSDB; AAS93634.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsive for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID NO 59806; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostic, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 705 AA;

Alignment Scores:

Pred. No.:	4e-30	Length:	705
Score:	405.50	Matches:	105
Percent Similarity:	37.384	Conservative:	9
Best Local Similarity:	34.434	Mismatches:	10
Query Match:	34.844	Indels:	181
DB:	22	Gaps:	7

US-09-809-545A-1_COPY_535_1143 (1-609) X ABG29447 (1-705)

QY	157	GTATATCTTCTGCAATGCTGCTTTCATCCGCGCCGCACTGCTGAGCTGATAC	216
DB	190	ValGluLeuSerValProGlyPheProTyrProAlaIleThrAlaAlaAlaTyr	209
QY	217	CGAGGGGCTCACTTCGAGGCGGTGTGCGACCGTGTACAACTTCAGAGCTGCGCG	276
DB	210	ArgGlyAlaIleIleuAlaArgGlyArgGlyArgThrValTyrAsnThrPheAlaIleAla	229
QY	277	CCCCCACCCTCAATCCCGGCTATGGCGAGTA	309
DB	230	ProProProIleProIleProAlaTyrGlyLeuThrPasnLeuValAspLeuProGluTyr	249
QY	310	-----GTGTAT-----	315
DB	250	PheAlaLeuSerAspAspAspPheAsnIleLeuLeuLeuValCysIleArgValTyrThr	269
QY	316	-----CAAGAG-----	321
DB	270	GlyGlnGlnGlnThrSerLeuCysHicCysProAlaGlnGlnGlnAsnLysArgAspLeu	289
QY	322	-----CAGTGTATGGC	333
DB	290	GlnLeuLysMetPheValAspAlaGlnCysProPheGlnLeuGlyGlyProVal-----	307
QY	334	AATAAATTGCTACAGCGTGTGTACGTCGATACCGCTTACCGCCGACCCCTGGCACT	393
DB	308	ProArgLeuGlyGlnGlyGlyTyrAlaIleTyrArgTyrAlaGlnProThrProAlaThr	327
QY	394	GCTGCTGCTTACAGTAC-----	411
DB	328	AlaAlaAlaTyrSerAspArgLeuSerLeuMetTyrGlyAspGlySerGluIleValArg	347
QY	411	-----	411
DB	348	LysGlyArgAsnGlnPheValPheValAlaAlaAspGluIleSerCysAsnThrSerAla	367
QY	411	-----	411
DB	368	GlyCysTyrProSerTyrGlnLeuGlnAlaSerSerGlyValGlnGlyValThrSer	387
QY	411	-----	411
DB	368	AlaPheIlePheSerIleProThrSerGlnProThrIleAlaIleSerTyrIleLeuThr	407
QY	411	-----	411
DB	408	LeuAspSerGlyAspAlaMetAlaAspLysIleAspThrAlaProAlaSerMetGluPhe	427
QY	412	-----AGTTACGAGACGAGTTTATGCTGCCGACCCCTTACACACACACA	453
DB	428	SerPheGlnGlnGlyGlySerTyrGlyArgValTyrAlaAlaAspProTyrIleIleAla	447
QY	454	CTTGCTCAGACCCCGACCTTACGAGCGGTGTGTCGACGAAATGCTTTGGCCCTTGACCAT	513
DB	448	LeuAlaProAlaIleuAlaArgGlyArgGlyArgThrValTyrAsnThrPheAlaIleAla	460
QY	514	GCCAAGACTAGGACCATGCTGATGATGTGGCTCTGTTCTTCATTCAGAGCTAGT	573
DB	461	-----ArgGlnTyr-----SerSerLeuCysValSer	469
QY	574	ATAATACCAAGGGGA	588
DB	470	ValCysThrGlyGly	474
RESULT 10			
AA41878			
ID	AA41878	standard; Protein: 291 AA.	
AC	AA41878;		
XX			
DT	22-OCT-2001	(first entry)	
XX			
DE	Human polypeptide SEQ ID NO 6809.		

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 268 AA;

SO Alignment Scores:

Pred. No.:	1,61e-42	Length:	268
Score:	534.00	Matches:	105
Percent Similarity:	67.68%	Conservative:	6
Best Local Similarity:	64.02%	Mismatches:	23
Query Match:	45.88%	Indels:	30
DB:	22	Gaps:	4

US-09-809-545a-1_COPY_535_1143 (1-609) x AM40092 (1-268)

QY 1 ATGACATATAAAGGCGGTGACCCCTACACACATGGCTGGAATTAATTCAGTTGTG 60
DB 134 MetThrAnlyslvstHnglylAsnProlyrThrAnnglyrPlvslvleuAsnProvalVal 153
QY 61 GCGCGGTCTACAGCCCGGACTTGTATGACGACGCGTGTGTGCGACGCCAACAG 120
DB 154 GlyAlaValTyrGlyProGluPheTyrAlaValThr----- 165
QY 121 GAGGATCTTCATGTACAGTGGCCCGGACGTTCACTGTATATCTTCGCAATGCTGGC 180
DB 166 -----Gly 166
QY 181 TTTCATATCCGCGCGCCATCTGTACAGCTGATACGAGGGGCTCACTTCAGGCGGT 240
DB 167 PheProlyrProthrThrnglylThrAlaValAlaTyrnglylAlaHsleuArgGlyArg 186
QY 241 GGTGCAACCGTGACACACCTTCAGAGCTGGCGCGCCCGCCCAATCCGCGCTAT 300
DB 187 GlyArgAlaValTyrAnthrPheArgAlaAlaProProProProleProThrTyr 206
QY 301 GCGGAGTACGTATTCAGAGCCAGTGTATGCAATAATGCTACAGGGGTGTTACGCT 360
DB 207 GlyAlaValAlaTyrGlnAspGlyPheTyrGlyAlaGlu---IleTyrGlyGlyTyrAla 225
QY 361 GCATACGCTACGCGCCAGCCCGGACCCCTGCTGCTGCTACAGTACAGTACGGA 420
DB 226 AlaTyrArgTyrAlaGlnProAla---AlaAlaAlaAlaAlaTyrSerAspSerTyrGly 244
QY 421 CGAGTTAT---GCTGCGAGCCCTTACACACACACTTGTCCAGCCCGCCACCTACGCG 477
DB 245 ArgAlaTyrAlaAlaAlaAspProTyrHisThrIleGlyProAlaAlaThrTyrSer 264
QY 478 GTTGTGTCATG 489
DB 265 IleGlyThrMet 268

RESULT 8
ABG29446
ID ABG29446 standard; Protein: 101 AA.
XX
AC ABG29446;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #29437.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
PN MO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI, 2001-639362/73.

PT N-PSDB; AAS93633.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID No 59805; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcc_sequences.

XX Sequence 101 AA;

Alignment Scores:			
Pred. No.:	2.86e-31	Length:	101
Score:	415.00	Matches:	84
Percent Similarity:	83.65%	Conservative:	3
Best Local Similarity:	80.77%	Mismatches:	13
Query Match:	35.65%	Indels:	4
DB:	22	Gaps:	2

US-09-809-545a-1_COPY_535_1143 (1-609) x ABG29446 (1-101)

QY 205 GCAGTGCATACCGAGGGGCTCACTTCGAGCGGTGTGCGACCGGTACACACCTTC 264
DB 1 AlaAlaAlaTyrnglylAlaHsleuArgGlyArgGlyArgThrAlaThrPhe 20
QY 265 AGAGTGGCGGCGCCCGCCCAATCCCGGCTATGCGGAGTGTATCAAGAGCCA 324
DB 21 ArgAlaAlaAlaProProProProleProAlaTyrGlyAlaValTyrGlnAspGly 40
QY 325 GTGTATGCGCAATAATGTCACAGGGTGTACGTCGATACCGGTACCGCCAGCCAC 384
DB 41 PheTyrGly-AlaAspIleTyrGlyAlaMet---LeuHisThrAlaThrProSerLeuP 59
QY 385 CTGCGACCTGCTGCTGCTACAGTACAGTACGAGGATTTATG---CTGCGGAGCC 441
DB 59 OlvsnProleuProleuProThrValThrValThrAspGluPheMetThrPhe-SerPro 79
QY 442 TACACACACACTTGTCTGAGCCCGCCACCTTACGAGGCTTGTGTCATGATGCTTTGCG 501
DB 79 euserProleuThrCySerSerProHisleuArgAlaGlyTyrGlnCySerPheCyt 99
QY 502 CCCTTGAC 509
DB 99 hrPheAsp 101

PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 XX WPI: 2001-442253/47.
 DR N-PSDB; AAI59249.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 5; SEQ ID NO 3238; 10078bp; English.
 XX
 XX The invention relates to human nucleic acids (AAI5798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 XX Sequence 237 AA;
 XX
 XX Alignment Scores:
 XX Pred. No.: 1,556-42 Length: 237
 XX Score: 534.00 Matches: 105
 XX Percent Similarity: 67.68% Conservative: 6
 XX Best Local Similarity: 64.02% Mismatches: 23
 XX Query Match: 45.88% Indels: 30
 XX DB: 22 Gaps: 4
 XX
 XX US-09-809-545A-1_COPY_535_1143 (1-609) x AAM40093 (1-237)
 QY 1 ATGACTAATATAAAGGCGCGTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
 DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 103 MetThrAsnIleValblyethrIglYsnProIYrThrAsnGlyTTrpIlyseuLenProValVal 122
 QY 61 GCGCGGCTTACAGCCCGCACTTCTATGAGGAGCGTGTGTTGCCAGGCCAACGAG 120
 DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 123 GAlaAlaValIYrGlyProGluPheYrAlaValThr----- 134
 DB 121 GAGGAGTCTTCATGATGACGTGGCCCGCACTTCTGTATATCTTCTGCAATGCTGGC 180
 QY |||||
 DB 135 -----Gly 135
 QY 181 TTTCATATCCGCGCGCACTGCTGAGCTGATGATCCGAGGGGCTGACCTGAGCGCGT 240
 DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 136 PheProIYrProIYrThrIglYThrAlaValAlaIYrArGlyAlaHisIeuArGlyArg 155
 QY 241 GGTGCAACCGTGTACAAACACCTTCAAGCTGGCGCGCCCGCCCACTCCCGGCTAT 300
 DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 156 GAlYArGAlValIYrAsnThrPheArGAlaAlaIProProIYrProIYr 175
 QY 301 GCGCGAGTGTATCAAGCGCAGAGTATGAGCAATAAATGTATACAGGGGTGTACGCT 360
 DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 176 GAlYAlaValIYrGlnAspGlyPheIYrGlyAlaGlu---IleIYrGlyGlyTYrAla 194
 QY 361 GCATACCGCTACGCCGACCCACCCCTGCGCATGCTGCTGCTTACAGTACAGTACGGA 420

DB 195 AlATyrArGTYrAlaGlnProAla---AlaAlaAlaAlaIATYrSerAspSerTYrGly 213
 QY 421 CGAGTTAT---GCTGCGCAGACCCCTACACCACTTCTCTCCAGCCCGCCACCTACGCG 477
 DB 214 ArGyAlATYrAlaAlaAlaAProIYrHisIeThrIleGlyProAlaAlaThrTYrSer 233
 QY 478 GTTGTCGCATG 489
 DB 234 IleGlyThrMet 237
 XX
 XX RESULT 7
 XX AAM40092
 XX ID AAM40092 standard; Protein: 268 AA.
 XX AAM40092;
 XX
 XX 22-OCT-2001 (first entry)
 XX
 XX Human polypeptide SEQ ID NO 3237.
 XX
 XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 XX peripheral nervous system; neuropathy; central nervous system; CNS;
 XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 XX leukaemia.
 XX
 XX Homo sapiens.
 XX
 XX WO200153312-A1.
 XX
 XX 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 XX
 XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX 09-JUL-2000; 2000US-0598042.
 XX 19-JUL-2000; 2000US-0620312.
 XX 03-AUG-2000; 2000US-0653450.
 XX 14-SEP-2000; 2000US-0662191.
 XX 19-OCT-2000; 2000US-0693036.
 XX 29-NOV-2000; 2000US-0727344.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 XX WPI: 2001-442253/47.
 XX N-PSDB; AAI59248.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 XX PT such as central nervous system injuries -
 XX
 XX PS Example 5; SEQ ID NO 3237; 10078bp; English.
 XX
 XX The invention relates to human nucleic acids (AAI5798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.


```
QY 61 GCGCGGTCTACAGCCCGCACTTATGACAGGACAGGTCCTTGTGCGACCAACGAG 120
DB 259 GAlaValValTyrSerProGluPheThrAlaGlyThrValLeuLeuCyGlnAlaIasnGln 278
QY 121 GAGGATCTTCATGATACAGTGGCCCACTTCACTTGATATACTTCTTGCAATGCTGGC 180
DB 279 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 298
QY 181 TTTCCATATCCGCGCCGCACTGCTGACGTGATACCGAGGGGCTCAGCTTGAGCGCT 240
DB 299 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 318
QY 241 GATGCGACCGGTATACACACCTTCAGAGCTGGCGGCCCGCCCAATCCCGGCTAT 300
DB 319 GlAArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 338
QY 301 GCGGAGTAGTGTATACAGACCGCATGTATGGCAATTAATTGCTACAGGGTGTACGCT 360
DB 339 GlYglYValValTyrGlnAspGlyPheTyrGlyAlaAsp---IleTyrGlyGlyTyrAla 357
QY 361 GCATCCGCTACGCCGACGCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 358 AlAArgTyrTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 377
QY 421 CGAGTTATGCTGCGACCGCCCTACACACACACACTTGTCTCCAGCCCGCACTACGGCGTT 480
DB 378 ArgValTyrAlaAlaAspProTyrHisAlaLeuAlaProAlaProThrTyrGlyVal 397
QY 481 GGTGCATGATGCTTTTGGCGCCTTGACCGATGCCAGACTAGAGACCATCTGATGAT 540
DB 398 GlAAlaMetAlaAlaPheAlaPheAlaProLeuThrAspAlaTyrThrArgSerHisAlaAsp 417
QY 541 GGGGCTCGCTCTTCTTCATGAGGCTATATATACCAAGGGGATACAAACGTTT 600
DB 418 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAsnArgPhe 437
QY 601 GCTCCATAT 609
DB 438 AlaProTyr 440

RESULT 4
AAY81462
ID AAY81462 standard; Protein: 330 AA.
XX
AC AAY81462;
XX
DT 03-JUN-2000 (first entry)
XX
DE Human ataxin-2 binding protein (A2BP).
XX
KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
KW neuronal plasticity; cellular degeneration signal transduction pathway;
KW selective RNA transport; spinocerebellar ataxia type-2;
KW hyperproliferative disorder.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1023..1312
FT notes="This region contains the regions necessary
FT for binding to ataxin-2"
XX
XX MO200012710-A1.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20156.
XX
XX 01-SEP-1998; 98US-0145391.
XX
XX (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
```

```
PI Puls SM, Shibata H;
XX
XX MPI: 2000-237873/20.
DR
DR N-PSDB: AAA07075.
XX
XX Nucleic acids encoding an ataxin-2 binding protein useful for
PT inhibiting the expression of active proteins from the SCA2 gene for the
PT treatment of spinocerebellar ataxia type-2 -
XX
XX Claim 19; Page 77-78; 82pp; English.
XX
XX This sequence represents cDNA encoding human ataxin-2 binding protein
XX (A2BP). Nucleotide sequences encoding human A2BP were originally
XX isolated in an adult brain cDNA library using the yeast two hybrid
XX method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
XX function that is encoded by the SCA2 gene located on chromosome 12. SCA2
XX has been linked to the autosomal dominant neurodegenerative disorder
XX spinocerebellar ataxia type-2. Individuals afflicted with the disease
XX exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
XX containing a polyglutamine stretch of about 35-39 residues, whereas that
XX of normal individuals contains approximately 22 contiguous glutamine
XX residues. A2BP and ataxin-2 are components of a cellular degeneration
XX signal transduction pathway. The pathogenic expanded form of ataxin-2 has
XX a higher affinity for A2BP relative to normal ataxin-2; the presence of
XX the expanded form is likely to promote degeneration. A2BP and ataxin have
XX also been found to have a role in gene regulation. The binding of A2BP to
XX ataxin-2 plays an important role in controlling gene expression via the
XX targeting of transport of specific RNAs, selective RNA transport being
XX mediated via the RNA binding domains of A2BP. A2BP is expressed very
XX early in embryonic development. Both ataxin-2 and A2BP are able to bind
XX CC RNA, and are essential components of the RNA localisation network that
XX establishes cellular polarity in embryogenesis. In highly differentiated,
XX CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
XX CC function and are required for neuronal plasticity. A2BP nucleic acids may
XX be used for the recombinant production of A2BP proteins or fragments
XX thereof according to standard methodologies. For example, an A2BP protein
XX CC with an ataxin-2 or RNA binding capability but no signal transduction
XX CC function can be used as a dominant negative inhibitor of the cellular
XX CC degeneration signal transduction pathway. A2BP proteins with a signal
XX CC transduction function can be used to treat hyperproliferative disorders
XX (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX
XX Sequence 330 AA;
XX
XX Alignment Scores:
XX Pred. No.: 4,28e-51 Length: 330
XX Score: 624.00 Matches: 119
XX Percent Similarity: 89.78% Conservative: 4
XX Best Local Similarity: 86.86% Mismatches: 12
XX Query Match: 53.61% Indels: 2
XX DB: 21 Gaps: 0
XX
XX US-09-809-545A-1_COPY_535_1143 (1-609) x AAY81462 (1-330)
QY 1 ATGACTAATAAAGCCGTGAACCCCTACACCAATGCTGGAATTAATCACTGTG 60
DB 196 MetThrAsnLysLysThrValAsnProTyrThrAsnGlyTyrPlyLeuAsnProValVal 215
QY 61 GCGCGGTCTACAGCCCGCACTTATGACAGGACAGGTCCTTGTGCGACCAACGAG 120
DB 216 GlAAlaValValTyrSerProGluPheThrAlaGlyThrValLeuLeuCyGlnAlaIasnGln 235
QY 121 GAGGATCTTCATGATACAGTGGCCCACTTCACTTGATATACTTCTTGCAATGCTGGC 180
DB 236 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255
QY 181 TTTCCATATCCGCGCCGCACTGCTGACGTGATACCGAGGGGCTCAGCTTGAGCGCT 240
DB 256 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 275
QY 241 GGTGCGACCGGTATACACACCTTCAGAGCTGGCGGCCCGCCCAATCCCGGCTAT 300
DB 276 GlAArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 295
```

CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 397 AA;

Alignment Scores:

Pred. No.:	6,58e-88	Length:	397
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	86.73%	Indels:	1
		Gaps:	1

US-09-809-545A-1_COPY_535_1143 (1-609) x AAB92573 (1-397)

OY 1 ATGACTATAATAAAGCGCGTGAACCCCTACACCAATGCGTGAATTAATCCAGTTGTG 60
DB 196 MetThrAsnYslySThrValAsnProTyrThrAsnGlyTrpIysLeuAsnProValVal 215
OY 61 GCGCGGTCTACAGCCCGGCACTTCTATGACGACGCGTGTGTCGACGACCAACAG 120
DB 216 G1yAlaValTyrSerProGluPheTyrAlaGlyThrValLeuLeuCyG1AlaAsnGln 235
OY 121 GAGGAGTCTTCAGATGATACAGTGGCCCGGCACTTCTATGATTAATCTTCGACGCGC 180
DB 236 GluIysSerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProIy 255
OY 181 TTTCATATCCGCGCGGCACTGCTGACACTGATACGACGAGGCGCTACCTTCGAGCGCCT 240
DB 256 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrAlaGlyAlaHisLeuArgGlyArg 275
OY 241 GGTGCAACCGTGTACACACCTTCAGACTGCGGCGGCGGCGGCGGCGGCGCTAT 300
DB 276 G1yArgThrValTyrAsnThrPheArgAlaAlaAlaProProProIleProAlaTyr 295
OY 301 GCGCGAGTACGTATCAAGACCGCACTGATGCAATTAATGCTCAAGCGTGTACGCT 360
DB 296 G1yG1yAlaValTyrGlnAspGlyPheTyrG1yAlaAsp---11eTyrG1yG1yTyrAla 314
OY 361 GCATACGCGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCT 420
DB 315 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 334
OY 421 CGAGTTATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 480
DB 335 ArgValTyrAlaAlaAspProTyrHisAlaLeuAlaProAlaProThrTyrGlyAla 354
OY 481 GGTGCGATGAATGCTTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 540
DB 355 G1yAlaMetAsnAlaPheAlaProLeuThrAspAlaTyrSerHisAlaAspAsp 374
OY 541 GTGGGTCCGTTCTTTCTTCATTCGAGGCTAGTATATCAAGAGGCGGTACACGCTTT 600
DB 375 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAsnArgPhe 394
OY 601 GCTCCATAT 609
DB 395 AlaProTyr 397

RESULT 3
AAM79112
ID AAM79112 standard; Protein: 440 AA.

XX XX
AC AAM79112;
XX 06-NOV-2001 (first entry)
DT XX
XX
DE Human protein SEQ ID NO 1774.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX MO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001MO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEO INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Pi Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAK52245.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX
PS Claim 20; Page 4126-4127; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78333-AAK80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activation/inhibit activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.

SQ Sequence 440 AA;

Alignment Scores:

Pred. No.:	6.77e-88	Length:	440
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	86.73%	Indels:	1
		Gaps:	1

US-09-809-545A-1_COPY_535_1143 (1-609) x AAM79112 (1-440)

OY 1 ATGACTATAATAAAGCGCGTGAACCCCTACACCAATGCGTGAATTAATCCAGTTGTG 60
DB 239 MetThrAsnYslySThrValAsnProTyrThrAsnGlyTrpIysLeuAsnProValVal 258

KM Alzheimer's disease; gene therapy.
 OS Rattus norvegicus.
 XX
 PN W0200174901-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09555.
 XX
 PR 31-MAR-2000; 2000US-193548P.
 PR 14-MAR-2001; 2001US-0809545.
 XX
 PA (SCIO-) SCIOS INC.
 XX
 PI Stanton LW, White RT;
 DR WPI; 2002-010779/01.
 DR N-PSDB; AAS94693.
 XX
 PT Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX
 PS Claim 18; Fig 1; 1899P; English.
 XX

CC The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAU70146-AAU70178 represent the secreted factor
 CC polypeptides of the invention.
 CC
 XX

XX Sequence 203 AA;

Alignment Scores:
 Pred. No.: 2,676-95 Length: 203
 Score: 1086.00 Matches: 203
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 93.30% Indels: 0
 DB: 23 Gaps: 0

US-09-809-545A-1_COPY_535_1143 (1-609) x AAU70146 (1-203)

QY 1 ATGACTAATATAAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTG 60
 DB 1 MetThrAsnIlybLysAlaValAsnProYrYrAsnGlyTTrpLysLeuAsnProValVal 20
 QY 61 GCGCGGCTACAGCCCGCACTTCTATGAGAGCGCTGCTGTGGCCAGGCACACAG 120
 DB 21 GYAlaValIlySerProAspPheYrAlaGlyThValLeuLeuYsgInAlaAsnGln 40
 QY 121 GAGGATCTTCATGATAGTGGCCCGCACTTCACTTGATATTAATTCCTCAATGCTGGC 180
 DB 41 GluGlySerSerMetYrSerGlyProSerSerLeuValIlyrThSerAlaMetProGly 60
 QY 181 TTTCATATCCGCGCGCACTGCTGAGCTGATACCGAGGGCTACCTTGGAGCGCGT 240
 DB 61 PheProIlyrProAlaAlaThrAlaAlaAlaIlyrArgGlyAlaHisLeuArgGlyArg 80
 QY 241 GGTCCGACCGGTGACAAACCTTCAGAGCTGGCGGCCCGCCACCCCATCCCGGCTAT 300
 DB 81 GYATrGThValIlyrAsnThrPheArgAlaAlaAlaProProProIleProAlaIlyr 100

QY 301 GCGGAGTAGTGTATCAAGACCCAGGTATGCGAATTAATTCCTACAGGGTTCACCT 360
 DB 101 GYGIyAlaValIlyrGlnGlnProValIlyrGIyAsnYsLeuLeuGlnGlyGIyAla 120
 QY 361 GCATACCGCTACGCCCGCACCCCGCCACCTGCTGCTGCTACAGTACAGTACAGCA 420
 DB 121 AlATyArGlyrAlaGlnProThrProAlaThrAlaAlaIlyrSerAspSerIly 140
 QY 421 CGAGTTATGCTGCCGACCCCTACACACACACACTTGTCTCAGCCCCCAGTACGGCGTT 480
 DB 141 ArgValIlyrAlaAlaAspProIlyrHisThrLeuAlaProAlaProThrIlyrGIyAl 160
 QY 481 GGTCCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATAT 540
 DB 161 GYAlaMetAsnAlaPheAlaProLeuThrAspAlaIlyrThrArgSerHisAlaAsp 180
 QY 541 GTGGGTCTCGCTTTCTTCTTCAATTCAGGCTGTATTAATCAAGGGGATACACCGTTT 600
 DB 181 ValGIyLeuValLeuSerSerLeuGlnAlaSerIlyrGlnGlyGIyTrAsnArgPhe 200
 QY 601 GCTCCATAT 609
 DB 201 AlaProIlyr 203

RESULT 2
 AAB92573
 ID AAB92573 standard; Protein; 397 AA.
 XX
 AC AAB92573;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:10787.
 XX
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUN-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Ito gai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugaiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 10787; 2537p + CD ROM; English.
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a

THIS PAGE BLANK (USPTO)

```

; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-482-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

RESULT 13
US-10-176-757-465
; Sequence 465, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-757-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

RESULT 14
US-10-176-913-465
; Sequence 465, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
US-10-176-913-465

```

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-913-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

RESULT 15
US-10-180-552-465
; Sequence 465, Application US/10180552
; Publication No. US20030022300A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C153
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-180-552-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

Search completed: March 16, 2003, 02:35:20
Job time : 80.3694 secs

```


LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-737-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 GCCCCACCCCAATCCC 293
Db 3096 GCCCCACCCCAATCCC 3113

RESULT 9

US-10-173-706-465
Sequence 465, Application US/10173706
Publication No. US20030022293A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C7
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-173-706-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 GCCCCACCCCAATCCC 293
Db 3096 GCCCCACCCCAATCCC 3113

RESULT 10

US-10-175-738-465
Sequence 465, Application US/10175738
Publication No. US20030022294A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C45
CURRENT FILING DATE: 2002-06-19

Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-738-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 GCCCCACCCCAATCCC 293
Db 3096 GCCCCACCCCAATCCC 3113

RESULT 11

US-10-175-752-465
Sequence 465, Application US/10175752
Publication No. US20030022295A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-752-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 GCCCCACCCCAATCCC 293
Db 3096 GCCCCACCCCAATCCC 3113

RESULT 12

US-10-176-482-465
Sequence 465, Application US/10176482
Publication No. US20030022296A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

ORGANISM: Homosapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: n=a,t,g or c
US-09-969-708-535

Query Match 3.0%; Score 18; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 CTGTCTCAGCCCCACC 471
DB 142 CTGTCTCAGCCCCACC 125

RESULT 5
US-10-102-558-1/c
Sequence 1, Application US/10102558
Publication No. US2003008706X1
GENERAL INFORMATION:
APPLICANT: Dai, Ken-Shwo
TITLE OF INVENTION: HUMAN PEN1B-RELATED GENE VARIANT ASSOCIATED WITH LUNG CANCERS
FILE REFERENCE: U 013922-3
CURRENT FILING DATE: 2002-03-20
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 2173
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (128)..(574)
OTHER INFORMATION:
US-10-102-558-1

Query Match 3.0%; Score 18; DB 9; Length 2173;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGGCGCCCCCACC 288
DB 1260 GCGGCGCCCCCACC 1243

RESULT 6
US-10-174-590-465
Sequence 465, Application US/10174590
Publication No. US2003008352A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C42
CURRENT FILING DATE: 2002-06-18
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien

US-10-174-590-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCGGCGCCCCCATCCC 293
DB 3096 GCGGCGCCCCCATCCC 3113

RESULT 7
US-10-176-758-465
Sequence 465, Application US/10176758
Publication No. US2003008353A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C104
CURRENT FILING DATE: 2002-06-21
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-758-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCGGCGCCCCCATCCC 293
DB 3096 GCGGCGCCCCCATCCC 3113

RESULT 8
US-10-175-737-465
Sequence 465, Application US/10175737
Publication No. US2003001315A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C50
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465

Qy	301	GGCGAGTAGTATCAAGAGCCAGTGTATGGCAATAAATGGTACAGGGTGGTAACCT	360
Db	835	GGCGAGTAGTATCAAGAGCCAGTGTATGGCAATAAATGGTACAGGGTGGTAACCT	894
Qy	361	GCATACCGCTACGCCCAGCCACCCTTCGCACCTGCTGCCTTACAGTACAGTTACGGA	420
Db	895	GCATACCGCTACGCCCAGCCACCCTTCGCACCTGCTGCCTTACAGTACAGTTACGGA	954
Qy	421	CGAGTTATGCTGCCGAGCCCTTACACACACACTTGTCCAGCCCCACCTTACGGCGTT	480
Db	955	CGAGTTATGCTGCCGAGCCCTTACACACACTTGTCTCAGCCCCACCTTACGGCGTT	1014
Qy	481	GGTGCATGAATGCTTTTGGGCCCTTACCGAGTCCCAAGTACGAGCCATCTGTATGAT	540
Db	1015	GGTGCATGAATGCTTTTGGGCCCTTACCGAGTCCCAAGTACGAGCCATCTGTATGAT	1074
Qy	541	GTGGGTCCGATCTTTCTTCTTATTCGACGGCTAGTATATACCAAGGGGATACACCGTTT	600
Db	1075	GTGGGTCCGATCTTTCTTCTTCTTATTCGACGGCTAGTATATACCAAGGGGATACACCGTTT	1133
Qy	601	GCTCCATAT 609	
Db	1135	GCTCCATAT 1143	

```

RESULT 2
US-09-794-591-1
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulet, Stefan M.
; APPLICANT: Shbate, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987)..(1979)
US-09-794-591-1

```

Query Match	Similarity	11.5%	Score 70	DB 10	Length 2372
Best Local	Similarity	100.0%	Pred. No. 6-28		
Matches	70	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
QY	511	GATGCCAAGACTGAGGCCATGCTGATGATGAGGAGCTGCTTCTTCTCATTCAGAGGCT	570		
Db	2133	GATGCCAAGACTGAGGCCATGCTGATGATGATGAGGAGCTGCTTCTTCTCATTCAGAGGCT	2192		
QY	571	AGTATATACC	580		
Db	2193	AGTATATACC	2202		

RESULT 3
 US-09-783-590-7497
 Sentence 7497, Application US/09783590
 Patent No. US2002010850A1
 GENERAL INFORMATION:
 APPLICANT: Dillon, Patrick J.
 APPLICANT: Hesseline, William A.
 APPLICANT: Li, Haodong
 APPLICANT: Rosen, Craig A.

```

1  APPLICANT: Ruben, Steven M.
2  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
3  FILE REFERENCE: PO-16.2C1
4  CURRENT APPLICATION NUMBER: US/09/783,590
5  CURRENT FILING DATE: 2000-02-15
6  PRIOR APPLICATION NUMBER: 08/420,856
7  PRIOR FILING DATE: 1995-04-12
8  PRIOR APPLICATION NUMBER: 08/346,731
9  PRIOR FILING DATE: 1994-11-21
10 NUMBER OF SEQ. ID NOS: 12485
11 SOFTWARE: PatentIn Ver. 2.0
12 SEQ. ID NO 7487
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
100
```

```

1 ORGANISM: Homo sapiens
2 FEATURE:
3 NAME/KEY: misc feature
4 LOCATION: (31)
5 OTHER INFORMATION: n equals a,t,g, or c
6 NAME/KEY: misc feature
7 LOCATION: (137)
8 OTHER INFORMATION: n equals a,t,g, or c
9 NAME/KEY: misc feature
10 LOCATION: (161)
11 OTHER INFORMATION: n equals a,t,g, or c
12 NAME/KEY: misc feature
13 LOCATION: (165)
14 OTHER INFORMATION: n equals a,t,g, or c
15 NAME/KEY: misc feature
16 LOCATION: (172)
17 OTHER INFORMATION: n equals a,t,g, or c
18 NAME/KEY: misc feature
19 LOCATION: (252)
20 OTHER INFORMATION: n equals a,t,g, or c
21 NAME/KEY: misc feature
22 LOCATION: (266)
23 OTHER INFORMATION: n equals a,t,g, or c
24 NAME/KEY: misc feature
25 LOCATION: (271)
26 OTHER INFORMATION: n equals a,t,g, or c
27
28 US-09-783-590-7497

```

Query	Match	Similarity	3.0%	Score	18	DB	10	Length	276	
	Best	Local	Similarity	100.0%	Pred.	No.	8.4			
	Matches	18	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	276	GGCCCCACCCCAATCCC	293							
DB	115	GGCCCCACCCCAATCCC	132							

```

RESULT 4
US-09-969-708-535/C
; Sequence 535, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 535
; LENGTH: 332
; TYPE: DNA
;

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 16:12:39 ; Search time 54.9694 Seconds
(Without alignments)
7861.696 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgacctataaaagccgt.....acaacgcttcgctccatcat 609

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 501302 seqs, 350932545 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Published Applications NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	609	100.0	1340	10 US-09-809-545a-1	Sequence 1, Appl1
2	70	11.5	2372	10 US-09-794-591-1	Sequence 1, Appl1
3	18	3.0	278	10 US-09-783-590-7497	Sequence 7497, App
4	18	3.0	332	10 US-09-969-708-535	Sequence 535, App
5	18	3.0	2173	9 US-10-102-558-1	Sequence 1, Appl1
6	18	3.0	3582	9 US-10-174-580-465	Sequence 465, App
7	18	3.0	3582	9 US-10-176-758-465	Sequence 465, App
8	18	3.0	3582	9 US-10-175-737-465	Sequence 465, App
9	18	3.0	3582	9 US-10-173-706-465	Sequence 465, App
10	18	3.0	3582	9 US-10-175-738-465	Sequence 465, App
11	18	3.0	3582	9 US-10-175-752-465	Sequence 465, App
12	18	3.0	3582	9 US-10-176-482-465	Sequence 465, App
13	18	3.0	3582	9 US-10-176-757-465	Sequence 465, App
14	18	3.0	3582	9 US-10-176-913-465	Sequence 465, App
15	18	3.0	3582	9 US-10-180-552-465	Sequence 465, App
16	18	3.0	3582	9 US-10-180-557-465	Sequence 465, App
17	18	3.0	3582	9 US-10-173-700-465	Sequence 465, App
18	18	3.0	3582	9 US-10-174-572-465	Sequence 465, App
19	18	3.0	3582	9 US-10-174-579-465	Sequence 465, App

20	18	3.0	3582	9 US-10-174-582-465	Sequence 465, App
21	18	3.0	3582	9 US-10-174-588-465	Sequence 465, App
22	18	3.0	3582	9 US-10-175-739-465	Sequence 465, App
23	18	3.0	3582	9 US-10-175-740-465	Sequence 465, App
24	18	3.0	3582	9 US-10-175-743-465	Sequence 465, App
25	18	3.0	3582	9 US-10-176-488-465	Sequence 465, App
26	18	3.0	3582	9 US-10-176-492-465	Sequence 465, App
27	18	3.0	3582	9 US-10-176-747-465	Sequence 465, App
28	18	3.0	3582	9 US-10-176-750-465	Sequence 465, App
29	18	3.0	3582	9 US-10-176-985-465	Sequence 465, App
30	18	3.0	3582	9 US-10-176-987-465	Sequence 465, App
31	18	3.0	3582	9 US-10-176-991-465	Sequence 465, App
32	18	3.0	3582	9 US-10-176-992-465	Sequence 465, App
33	18	3.0	3582	9 US-10-176-993-465	Sequence 465, App
34	18	3.0	3582	9 US-10-184-658-465	Sequence 465, App
35	18	3.0	3582	9 US-10-173-695-465	Sequence 465, App
36	18	3.0	3582	9 US-10-173-697-465	Sequence 465, App
37	18	3.0	3582	9 US-10-173-705-465	Sequence 465, App
38	18	3.0	3582	9 US-10-174-576-465	Sequence 465, App
39	18	3.0	3582	9 US-10-174-585-465	Sequence 465, App
40	18	3.0	3582	9 US-10-174-586-465	Sequence 465, App
41	18	3.0	3582	9 US-10-175-747-465	Sequence 465, App
42	18	3.0	3582	9 US-10-176-481-465	Sequence 465, App
43	18	3.0	3582	9 US-10-176-485-465	Sequence 465, App
44	18	3.0	3582	9 US-10-176-487-465	Sequence 465, App
45	18	3.0	3582	9 US-10-176-493-465	Sequence 465, App

ALIGNMENTS

RESULT 1
US-09-809-545a-1
Sequence 1, Application US/0909545A
Patent No. US20020110804A1
GENERAL INFORMATION:
APPLICANT: Stanton, Lawrence W.
APPLICANT: White, R. Tyler
TITLE OF INVENTION: SECRETED FACTORS
FILE REFERENCE: SCIOS.017A
CURRENT APPLICATION NUMBER: US/09/809,545A
CURRENT FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1340
TYPE: DNA
ORGANISM: Rattus norvegicus
US-09-809-545a-1

Query Match 100.0%; Score 609; DB 10; Length 1340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGACTAATATAAAGCGCGTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTG	60
Db	535	ATGACTAATATAAAGCGCGTGAAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG	594
Qy	61	GGCGCGGTACAGCCCGGACTTCTATGACGACGCGGTGTTGCGACGACCAACAG	120
Db	595	GGCGCGGTACAGCCCGGACTTCTATGACGACGCGGTGTTGCGACGACCAACAG	654
Qy	121	GAGGATCTTCATGATGACAGTGGCCCAAGTTCACTTGTATATACCTTGCAATCGCTGC	180
Db	655	GAGGATCTTCATGATGACAGTGGCCCAAGTTCACTTGTATATACCTTGCAATCGCTGC	714
Qy	181	TTTCCATATCCGCGCCCATCTGTCAGCTGCATATCCGAGGCGCTCACTTCGAGCGCT	240
Db	715	TTTCCATATCCGCGCCCATCTGTCAGCTGCATATCCGAGGCGCTCACTTCGAGCGCT	774
Qy	241	GGTGCACCGGTATCAACACTTCAAGAGCTCGGCGCCCAACCCCAATCCGCGCTAT	300
Db	775	GGTGCACCGGTATCAACACTTCAAGAGCTCGGCGCCCAACCCCAATCCGCGCTAT	834

THIS PAGE BLANK (USPTO)

STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578,615A
FILING DATE: 11-JAN-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852
FILING DATE: 16-MAR-1992
APPLICATION NUMBER: 08/089,996
FILING DATE: 09-JUL-1993
APPLICATION NUMBER: 08/199,779
FILING DATE: 22-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Paul K. Leggaard
REGISTRATION NUMBER: 38,534
REFERENCE/DOCKET NUMBER: ISIS-1568
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-578-615A-104

Query Match 2.8%; Score 17; DB 3; Length 1812;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 277 CCCCCCCCCCAATCCC 293
Db 328 CCCCCCCCCCAATCCC 312

RESULT 14
PCT-US94-07770-104/C
Sequence 104, Application PC/TUS9407770
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and
APPLICANT: Russell T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07770
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852

FILING DATE: March 16, 1992
APPLICATION NUMBER: 08/089,996
FILING DATE: July 9, 1993
APPLICATION NUMBER: 08/199,779
FILING DATE: February 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumont
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1546
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US94-07770-104

Query Match 2.8%; Score 17; DB 5; Length 1812;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 277 CCCCCCCCCCAATCCC 293
Db 328 CCCCCCCCCCAATCCC 312

RESULT 15
US-09-146-053-6
Sequence 6, Application US/09146053A
Patent No. 6399349
GENERAL INFORMATION:
APPLICANT: Ryan, James W.
APPLICANT: Sprinkle, Terry Joe Curtis
APPLICANT: Venema, Richard C.
TITLE OF INVENTION: Human Aminopeptidase P Gene
FILE REFERENCE: MCG103
CURRENT APPLICATION NUMBER: US/09/146,053A
CURRENT FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/057,854
EARLIER FILING DATE: 1997-09-02
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 45546
TYPE: DNA
ORGANISM: Homo sapiens
US-09-146-053-6

Query Match 2.8%; Score 17; DB 4; Length 45546;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 365 ACCGTAAGCCAGCC 381
Db 22991 ACCGTAAGCCAGCC 23007

Search completed: March 16, 2003, 02:32:35
Job time : 69.3715 secs

TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 1757 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-578-615A-105

Query Match 2.8%; Score 17; DB 3; Length 1757;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 CCCCCCCCCCAATCCC 293
Db 273 CCCCCCCCCCAATCCC 257

RESULT 11

PCT-US94-07770-105/C
Sequence 105, Application PC/TUS9407770
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and
Applicant: Russell T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of
NUMBER OF SEQUENCES: 119
KINASE C

CORRESPONDENCE ADDRESS:
ADDRESSER: Woodcock Washburn Kurtz
ADDRESSEE: Mackiewicz & Norris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07770
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852
FILING DATE: March 16, 1992
APPLICATION NUMBER: 08/089,996
FILING DATE: July 9, 1993
APPLICATION NUMBER: 08/189,779
FILING DATE: February 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumond
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1546
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3439
TELEFAX: (215) 568-3100
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 1757 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US94-07770-105

Query Match 2.8%; Score 17; DB 5; Length 1757;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 CCCCCCCCCCAATCCC 293
Db 273 CCCCCCCCCCAATCCC 257

RESULT 12

US-08-199-779-1/C
Sequence 1, Application US/08199779
Patent No. 5681747
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett
TITLE OF INVENTION: Oligonucleotide Modulation of
NUMBER OF SEQUENCES: 17
KINASE C

CORRESPONDENCE ADDRESS:
ADDRESSER: Woodcock Washburn Kurtz
ADDRESSEE: Mackiewicz & No. 5681747ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/199,779
FILING DATE: herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 089,996
FILING DATE: July 9, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumond
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-199-779-1

Query Match 2.8%; Score 17; DB 1; Length 1812;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 CCCCCCCCCCAATCCC 293
Db 328 CCCCCCCCCCAATCCC 312

RESULT 13

US-08-578-615A-104/C
Sequence 104, Application US/08578615A
Patent No. 6015892
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and Russell, T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of Protein Kinase C
NUMBER OF SEQUENCES: 122
CORRESPONDENCE ADDRESS:
ADDRESSER: Woodcock Washburn Kurtz Mackiewicz & No. 6015892ris LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia

APPLICANT: Rhee, Sang-Ki
APPLICANT: Lee, Eun-Hae
TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes
TITLE OF INVENTION: and Methods of Use Thereof
FILE REFERENCE: 1533 0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 7
LENGTH: 4830
TYPE: DNA
ORGANISM: Glucobacter suboxydans
US-09-296-284-7

Query Match 3.0%; Score 18; DB 4; Length 4830;
Best Local Similarity 100.0%; Pred. No. 6,7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 GCGCCCCCAGCCCAATC 291
DB 97 GCGCCCCCAGCCCAATC 114

RESULT 8
US-09-337-307A-10/c
Sequence 10, Application US/09337307A
Patent No. 6432692
GENERAL INFORMATION:
APPLICANT: Bradfield, Christopher A.
APPLICANT: Carver, Lucy A.
APPLICANT: Dunham, Elizabeth E.
TITLE OF INVENTION: Sensitive Biossay For Detecting Agonists Of The Aryl Hydrocarbon
FILE REFERENCE: MARF0105
CURRENT APPLICATION NUMBER: US/09/337,307A
CURRENT FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 10
LENGTH: 36
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-337-307A-10

Query Match 2.8%; Score 17; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 CCACCTGCTGACCTGCA 213
DB 30 CCACCTGCTGACCTGCA 14

RESULT 9
US-08-199-779-2/c
Sequence 2, Application US/08199779
Patent No. 5681747
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett
TITLE OF INVENTION: Oligonucleotide Modulation of
TITLE OF INVENTION: Protein Kinase C
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz
ADDRESSEE: Mackiewicz & No. 5681747is
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/199,779
FILING DATE: herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 089,996
FILING DATE: July 9, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumond
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1757 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-199-779-2

Query Match 2.8%; Score 17; DB 1; Length 1757;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 CCCCCCCCCCAATCCC 293
DB 273 CCCCCCCCCCAATCCC 257

RESULT 10
US-08-578-615A-105/c
Sequence 105, Application US/08578615A
Patent No. 6015892
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and Russell, T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of Protein Kinase C
NUMBER OF SEQUENCES: 122
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6015892is LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578,615A
FILING DATE: 11-JAN-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852
FILING DATE: 16-MAR-1992
APPLICATION NUMBER: 08/089,996
FILING DATE: 09-JUL-1993
APPLICATION NUMBER: 08/199,779
FILING DATE: 22-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Paul K. Leggaard
REGISTRATION NUMBER: 38,534
REFERENCE/DOCKET NUMBER: ISIS-1568
TELECOMMUNICATION INFORMATION:

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 29
LENGTH: 615
TYPE: DNA
ORGANISM: Gluconobacter suboxydans
US-09-296-284-29

Query Match 3.0%; Score 18; DB 4; Length 615;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 GCGCCCCCACCCTCATC 291
|||
Db 48 GCGCCCCCACCCTCATC 65

RESULT 3
US-09-296-284-28
Sequence 28, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:

APPLICANT: Choi, Eun-Sung
APPLICANT: Rhee, Sang-Ki
APPLICANT: Lee, Eun-Hae
TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
FILE REFERENCE: 1533.0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 28
LENGTH: 664
TYPE: DNA
ORGANISM: Gluconobacter suboxydans
US-09-296-284-28

Query Match 3.0%; Score 18; DB 4; Length 664;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 GCGCCCCCACCCTCATC 291
|||
Db 97 GCGCCCCCACCCTCATC 114

RESULT 4
US-09-930-181-3/c
Sequence 3, Application US/09930181
Patent No. 6455292
GENERAL INFORMATION:

APPLICANT: Origene Technologies
TITLE OF INVENTION: Full-Length Serine Protein Kinase in Brain and Pancreas
FILE REFERENCE: 16U 101 V1
CURRENT APPLICATION NUMBER: US/09/930,181
CURRENT FILING DATE: 2001-08-16
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Version 3.0
SEQ ID NO 3
LENGTH: 3364
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (482) .. (2239)
US-09-930-181-3

Query Match 3.0%; Score 18; DB 4; Length 3364;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGCGCCCCCACCCTCA 288
|||

Db 3039 GCGCGCCCCCACCCTCA 3022

RESULT 5
US-09-632-098-1
Sequence 1, Application US/09632098
Patent No. 6420154
GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.
APPLICANT: Baidur, Nand
APPLICANT: Bishop, Paul D.
TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
FILE REFERENCE: 99-39
CURRENT APPLICATION NUMBER: US/09/632,098
CURRENT FILING DATE: 2000-08-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 3431
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (37) ... (2442)
US-09-632-098-1

Query Match 3.0%; Score 18; DB 4; Length 3431;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCGCCCCCACCCTCATCC 293
|||
Db 2978 GCGCCCCCACCCTCATCC 2995

RESULT 6
US-09-632-098-3
Sequence 3, Application US/09632098
Patent No. 6420154
GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.
APPLICANT: Baidur, Nand
APPLICANT: Bishop, Paul D.
TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
FILE REFERENCE: 99-39
CURRENT APPLICATION NUMBER: US/09/632,098
CURRENT FILING DATE: 2000-08-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 3468
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (37) ... (2472)
US-09-632-098-3

Query Match 3.0%; Score 18; DB 4; Length 3468;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCGCCCCCACCCTCATCC 293
|||
Db 3015 GCGCCCCCACCCTCATCC 3032

RESULT 7
US-09-296-284-7
Sequence 7, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:
APPLICANT: Choi, Eun-Sung

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 16:12:15 ; Search time 34.3715 Seconds
(without alignments)
5433.755 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143
Perfect score: 609
Sequence: 1 atgactaataaaaagcgcgt.....acaacgcttcggtccatat 609

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	11.5	2372	4 US-09-145-391-1	Sequence 29, Appl
2	18	3.0	615	4 US-09-296-284-29	Sequence 28, Appl
3	18	3.0	664	4 US-09-296-284-28	Sequence 1, Appl
4	18	3.0	3364	4 US-09-930-181-3	Sequence 3, Appl
5	18	3.0	3431	4 US-09-633-098-1	Sequence 3, Appl
6	18	3.0	3468	4 US-09-633-098-3	Sequence 7, Appl
7	18	3.0	4830	4 US-09-296-284-7	Sequence 10, Appl
8	17	2.8	36	4 US-09-337-307A-10	Sequence 2, Appl
9	17	2.8	1757	1 US-08-199-779-2	Sequence 105, App
10	17	2.8	1757	3 US-08-578-615A-105	Sequence 1, Appl
11	17	2.8	1757	5 PCT-US94-07770-105	Sequence 1, Appl
12	17	2.8	1812	1 US-08-199-779-1	Sequence 104, App
13	17	2.8	1812	3 US-08-578-615A-104	Sequence 104, App
14	17	2.8	1812	5 PCT-US94-07770-104	Sequence 104, App
15	17	2.8	45546	4 US-09-146-053-6	Sequence 3, Appl
16	16	2.6	437	1 US-08-631-200-3	Sequence 4, Appl
17	16	2.6	437	1 US-08-631-200-4	Sequence 5, Appl
18	16	2.6	437	1 US-08-631-200-5	Sequence 3, Appl
19	16	2.6	437	1 US-08-829-553-3	Sequence 3, Appl
20	16	2.6	437	1 US-08-829-553-4	Sequence 5, Appl
21	16	2.6	437	1 US-08-829-553-5	Sequence 3, Appl
22	16	2.6	437	2 US-08-922-267A-3	Sequence 3, Appl
23	16	2.6	437	2 US-08-922-267A-4	Sequence 3, Appl
24	16	2.6	437	2 US-08-922-267A-5	Sequence 3, Appl
25	16	2.6	437	2 US-08-936-707A-3	Sequence 4, Appl
26	16	2.6	437	2 US-08-936-707A-4	Sequence 4, Appl
27	16	2.6	437	2 US-08-936-707A-5	Sequence 5, Appl

28	16	2.6	437	2 US-08-936-706A-3	Sequence 3, Appl
29	16	2.6	437	2 US-08-936-706A-4	Sequence 4, Appl
30	16	2.6	437	2 US-08-936-706A-5	Sequence 5, Appl
31	16	2.6	437	3 US-09-248-203-3	Sequence 3, Appl
32	16	2.6	437	3 US-09-248-203-4	Sequence 4, Appl
33	16	2.6	437	3 US-09-248-203-5	Sequence 5, Appl
34	16	2.6	437	4 US-09-406-071-3	Sequence 3, Appl
35	16	2.6	437	4 US-09-406-071-4	Sequence 4, Appl
36	16	2.6	437	4 US-09-406-071-5	Sequence 5, Appl
37	16	2.6	437	4 US-08-905-223-86	Sequence 86, Appl
38	16	2.6	480	1 US-08-630-592-5	Sequence 5, Appl
39	16	2.6	480	1 US-08-714-991-5	Sequence 5, Appl
40	16	2.6	480	2 US-08-837-029-1	Sequence 1, Appl
41	16	2.6	833	2 US-08-883-526-2	Sequence 1, Appl
42	16	2.6	1366	3 US-08-883-526-2	Sequence 1, Appl
43	16	2.6	1458	4 US-09-291-023A-1	Sequence 1, Appl
44	16	2.6	1605	2 US-09-000-846-1	Sequence 1, Appl
45	16	2.6	1799	4 US-09-810-347-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-145-391-1
Sequence 1, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulist, Stefan M.
TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145,391
CURRENT FILING DATE: 1998-09-01
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 2372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (987)..(1979)
US-09-145-391-1

Query Match 11.5%; Score 70; DB 4; Length 2372;

Best Local Similarity 100.0%; Pred. No. 6e-27; Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATCGCAAGACTAGAGCATGCTGATGATGTGGCTCTGTTCTTTCATTGCAAGGCT 570
Db 2133 GATCGCAAGACTAGAGCATGCTGATGATGTGGCTCTGTTCTTTCATTGCAAGGCT 2192

Qy 571 AGTATATACC 580
Db 2193 AGTATATACC 2202

RESULT 2
US-09-296-284-29
Sequence 29, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:
APPLICANT: Choi, Eun-Sung
APPLICANT: Rhee, Sang-Ki
TITLE OF INVENTION: Glucosyltransferase Suboxydases Sorbitol Dehydrogenase, Genes
FILE REFERENCE: 1533.0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87

oligo (dT) primer [5',
AATCGAGAAATTCGCGCCCTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT3 vector. Library
went through one round of normalization. Library
constructed by M. Falcina Bonaldo."

BASE COUNT	69 a	122 c	102 g	90 t
ORIGIN				

Query Match	11.5%	Score 70;	DB 9;	Length 383;
Best Local Similarity	100.0%	Pred. No. 1.6e-25;		
Matches 70;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 511 GATGCCAAGACTAGAGACCATGTCGATGATGTGGTCTCGTTCTTCTTCAATTGCAGGCT 570

Db 295 GATGCCAAGACTAGAGACCATGTCGATGATGTGGTCTCGTTCTTCTTCAATTGCAGGCT 354

QY	571	AGTATATACC	580
Db	355	AGTATATACC	364

RESULT 15

LOCUS	AI186273	391 bp	mRNA	linear	EST 28-OCT-1999
DEFINITION	qcd20b08.x1 Soares_placenta_8tc5weeks_2NBHD8tc09W Homo sapiens cDNA				

clone IMAGE:1724247 3', mRNA sequence

VERSION AI186273.1 GI:3736911

KEYWORDS

ORGANISM Homo sapiens

REFERENCE

TITLE

TITLE

JOURNAL

COMMENT

COMMENT
Contact: Robert Strausberg, Ph.D.

This clone is available royalty free.

Insert Length: 1800 Std Error: 0.00

High quality sequence stop: 369.

Source: U.S. Census Bureau, *Current Population Reports*, 1990.

son

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:172424"
/clone_1b="Soares placenta_8to9weeks_2NBH28to9W"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5
TGTTCACATCTGAGGTGGAGCGGCCGCACATTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). library constructed by Bento Soares and
M.Fatima Ronaldo."

```

BASE COUNT	73 a	129 c	107 g	82 t
ORIGIN				

Query match	11.5%	Score 70;	DB 9;	Length 391;
-------------	-------	-----------	-------	-------------

```
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

QY 511 GATGCCAAGACTAGGAGCCATGCTGATGATGTGGTCTCGTTCCTTCATTCAGGCT 570

Db	367	GATGCCAAGACTAGAGCCAGTCGATGATGATGGGTCTCGTTCTTTCTTCAATGCAGGCT	366
Qy	571	AGTATATACC	580
Db	367	AGTATATACC	376

```
Search completed: March 16, 2003, 02:29:44
Job time : 860.787 secs
```

Qy 571 AGTATATACC 580
|||||
Db 158 AGTATATACC 167

RESULT 12
D60520/c 363 bp mRNA linear EST 28-AUG-1995
LOCUS
DEFINITION HM114E06A Clontech human fetal brain polyA+ mRNA (#6535) Homo
D60520 sapiens cDNA clone GEN-114E06 3', mRNA sequence.
ACCESSION
VERSION D60520.1 GI:962159
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Fujitawa, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takai, H.,
A., Takeda, S., Matenabe, T., Takahashi, E., Hirai, Y., Maekawa, H.,
Shin, S. and Nakamura, Y.
Fujitawa et al. (1995)
Unpublished (1995)
COMMENT Contact: Tsutomu Fujitawa
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawanuchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.

FEATURES
Source
Location/Qualifiers
1..363
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-114E06"
/clone_1lb="Clontech human fetal brain polyA+ mRNA (#6535)"
/note="Male adult, hematopoietic tissue, stem cell"

BASE COUNT 90 a 84 c 97 g 87 t 5 others
ORIGIN

Query Match 11.5%; Score 70; DB 14; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.5e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGCGATGCTGATGCGGCTCGTCTTCTTCATTCAGAGCT 570
|||||
Db 201 GATGCCAAGACTAGAGCGATGCTGATGCGGCTCGTCTTCTTCATTCAGAGCT 142
|||||

Qy 571 AGTATATACC 580
|||||
Db 141 AGTATATACC 132

RESULT 13
A1095813 367 bp mRNA linear EST 05-OCT-1998
LOCUS
DEFINITION qb20g11.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1696868 3', mRNA sequence.
ACCESSION
VERSION A1095813
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov

FEATURES
Source
Location/Qualifiers
1..367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1696868"
/clone_1lb="Soares_pregnant_uterus_NbHPU"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pTTT3-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
ACCTGAGAGATTGCGGCGCCCTTTTCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTTT3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 65 a 116 c 96 g 90 t
ORIGIN

Query Match 11.5%; Score 70; DB 9; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.5e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGCGATGCTGATGCGGCTCGTCTTCTTCATTCAGAGCT 570
|||||
Db 288 GATGCCAAGACTAGAGCGATGCTGATGCGGCTCGTCTTCTTCATTCAGAGCT 347
|||||

Qy 571 AGTATATACC 580
|||||
Db 348 AGTATATACC 357

RESULT 14
A1189407 383 bp mRNA linear EST 28-OCT-1998
LOCUS
DEFINITION qd01n05.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1172489 3', mRNA sequence.
ACCESSION
VERSION A1189407
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 486 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 381.
Location/Qualifiers
1..383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1722489"
/clone_1lb="Soares_pregnant_uterus_NbHPU"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pTTT3-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -

Seq primer: M13/PUC Reverse.
 FEATURES
 source
 Location/Qualifiers
 1..408
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="UTSW_S385"
 /clone_1ib="UTSW Adult Mouse Skeletal Muscle Library"
 /sex="Pooled"
 /tissue_type="Diaphragm/Hind limb muscles"
 /cell_type="Skeletal muscle"
 /dev_stage="2 months"
 /lab_host="DH5a"
 /note="Vector: PAMPI0 (Gibco); Cloned unidirectionally. Primer: Oligo dt. RNA Isolation: cytoplasmic RNA preps (Manniacis); Cloning Technique: CNA Cloning (Clontech), Life Technologies; Average insert size: 1.8 Kb; Insertion site: TACGTCACCTGATTCGAGC--->. Other information regarding entire library may be found at http://pqa.smed.edu/Data/Libraries/microarray_cdna_librar ies.htm."

BASE COUNT 108 a 104 c 91 g 101 t 4 others
 ORIGIN

Query Match 13.1%; Score 80; DB 12; Length 403;
 Best Local Similarity 100.0%; Pred. No. 8.9e-31;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AACCCCTACCAATGCGGAAATTAATCCAGTGTGGCGCGGTCTACAGCCCGAC 81
 Db 232 AACCCCTACCAATGCGGAAATTAATCCAGTGTGGCGCGGTCTACAGCCCGAC 291

Qy 82 TTCTATGACGACGCGTCT 101
 Db 292 TTCTATGACGACGCGTCT 311

RESULT 10 228 bp mRNA linear EST 14-DEC-1993
 LOCUS Z28900/c
 DEFINITION HSB87F031 STRATAGENE Human skeletal muscle cDNA library, cat.
 #938215. Homo sapiens cDNA clone B7F03, mRNA sequence.
 ACCESSION Z28900
 VERSION Z28900.1 GI:434570
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 228)
 Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houllgate,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y., Sebasteian,Kabakchis,C. and Tessier,A.
 IMAGE: molecular integration of the analysis of the human genome and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 9527534
 JOURNAL COMMENT
 MEDLINE Contact: Genethon
 GENEBASE Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 strand(-), single read.
 Location/Qualifiers
 1..228
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="B7F03"
 /clone_1ib="STRATAGENE Human skeletal muscle cDNA library, cat. #938215."
 /sex="Female"

/tissue_type="skeletal muscle"
 /dev_stage="19 years"
 /note="Organ: leg muscle; Vector: Lambda ZAPII; Tissue from female, 19 years old, normal leg muscle. Cloning vector is Lambda ZAPII, in vivo excision from Lambda ZAPII to pBluescript SK(+). Genexpress library reference is B. "
 BASE COUNT 64 a 51 c 48 g 64 t 1 others
 ORIGIN

Query Match 11.5%; Score 70; DB 14; Length 228;
 Best Local Similarity 100.0%; Pred. No. 1.3e-25;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGTCTGTTCTTTTCATTGACGCT 570
 Db 201 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGTCTGTTCTTTTCATTGACGCT 142

Qy 571 AGTATATACC 580
 Db 141 AGTATATACC 132

RESULT 11 327 bp mRNA linear EST 21-APR-2001
 LOCUS BG186689
 DEFINITION RST5664 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG186689
 VERSION BG186689.1 GI:13708376
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 327)
 Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J., Lerner,U., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,E., Veloso,N., Kijka,A., Hess,J., Cothren,K., Lo,K., Offendacher,J., Danzig,J. and Ducar,M.
 Creation of genome-wide protein expression libraries using random activation of gene expression
 Nat. Biotechnol. 19 (5), 440-445 (2001)
 21227151
 JOURNAL COMMENT
 MEDLINE Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scai@athersys.com
 High quality sequence stop. 327.
 Location/Qualifiers
 1..327
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1ib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

FEATURES
 source
 Location/Qualifiers
 1..327
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1ib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 99 a 78 c 69 g 81 t
 ORIGIN

Query Match 11.5%; Score 70; DB 12; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.5e-25;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGTCTGTTCTTTTCATTGACGCT 570
 Db 98 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGTCTGTTCTTTTCATTGACGCT 157

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 538)
 Bernaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 9704447
 On Feb 17, 1998 this sequence version replaced gi:2888390.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@iuii.wiwiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult brain library. cDNA library preparation: M. Fatima Bernaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics This clone is also available through the I.M.A.G.E.
 Consortium at LNL (info@image.lnl.gov). IMAGE ID=1767889 The
 following repetitive elements were found in this cDNA sequence:
 105-154, >(CAA)n#Simple_repeat
 Seq primer: M13 Forward
 POLYA=NO.

FEATURES
 SOURCE
 1. 538
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-A0-ar-b-11-0-UI"
 /clone_lib="UI-R-A0"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; This library
 consists of a mixture of individually tagged normalized
 libraries constructed from rat placenta, adult lung, brain
 , liver, kidney, heart, spleen, ovary, and muscle. The tag
 is a string of 3-5 nucleotides present between the Not I
 site and the oligo-dT track which allows identification of
 the library of origin of a clone within the mixture."

BASE COUNT 146 a 115 c 102 g 175 t
 ORIGIN

Query Match 15.1%; Score 92; DB 9; Length 538;
 Best Local Similarity 100.0%; Pred. No. 4.9e-37;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 489 GAATGCTTTGGCCCTTGACGAGTCCAGACTAGGAGCCATGCTGATGGGCTCT 548
 Db 284 GAATGCTTTGGCCCTTGACGAGTCCAGACTAGGAGCCATGCTGATGGGCTCT 225

Cy 549 CGTCTTTCTTCATTGACGAGCTAGTATATACC 580
 Db 224 CTTCTTTCTTCATTGACGAGCTAGTATATACC 193

RESULT 8
 B1988957 600 bp mRNA linear EST 20-DEC-2001
 LOCUS 4023-45 Mouse E14.5 retina lambda ZAP II library Mus musculus cDNA,
 mRNA sequence.
 ACCESSION B1988957
 VERSION B1988957.1 GI:17959947
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Mu, X., Zhao, S., Pereshad, R., Heieh, T.-F., Scarpa, A., Wang, S.W.,
 White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
 TITLE Gene expression in the developing mouse retina by EST sequencing
 JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
 MEDLINE 21671825
 COMMENT Contact: Klein, WH
 Department of Biochemistry and Molecular Biology
 University of Texas M.D. Anderson Cancer Center
 Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
 Tel: 713 792 3646
 Fax: 713 790 0329.

FEATURES
 SOURCE
 Location/Qualifiers
 1. 600
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="Mouse E14.5 retina lambda ZAP II library"
 /tissue_type="neural retina"
 /dev_stage="embryonic day 14.5 post-fertilization"
 /note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
 Primer: Oligo dT. RNA isolation: cytoplasmic RNA preps
 (Mannatis); Cloning Technique: CVA Cloning (Clontech),
 Life Technologies); Average insert size: 1.8 Kb;
 Insertion site: TACGTCACGATTCGAGTGC--->. Other
 information regarding entire library may be found at
 http://pga.swmed.edu/Data/libraries/microarray_cdna_librar
 ies.htm."

BASE COUNT 171 a 124 c 122 g 183 t
 ORIGIN

Query Match 15.1%; Score 92; DB 13; Length 600;
 Best Local Similarity 100.0%; Pred. No. 5e-37;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 489 GAATGCTTTGGCCCTTGACGAGTCCAGACTAGGAGCCATGCTGATGGGCTCT 548
 Db 450 GAATGCTTTGGCCCTTGACGAGTCCAGACTAGGAGCCATGCTGATGGGCTCT 509

Cy 549 CGTCTTTCTTCATTGACGAGCTAGTATATACC 580
 Db 510 CTTCTTTCTTCATTGACGAGCTAGTATATACC 541

RESULT 9
 BG795922 408 bp mRNA linear EST 16-MAY-2001
 LOCUS UTSW_SM38B5 UTSW Adult Mouse Skeletal Muscle Library Mus musculus
 CDNA clone UTSW_SM38B5, mRNA sequence.
 ACCESSION BG795922
 VERSION BG795922.1 GI:14131492
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 408)
 Gallardo, T.D., Schageman, J.J., Pertsemildis, A., Garner, H.R.,
 Williams, R.S. and Shohet, R.V.
 UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA
 Library
 Unpublished (2001)
 Contact: Schageman, JJ
 Shohet/Garner Labs
 University of Texas Southwestern Medical Center
 6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
 Tel: 214 648 1674
 Email: Jeff.Schageman@UTSouthwestern.edu
 cDNA library constructed by UTSW and the Reynolds Heart Disease
 Genomic Applications (PGA) and the Reynolds Heart Disease
 Prevention grants for use in cDNA microarray experiments. Sequence
 Quality: Sequence ends were trimmed based on percentage of ambigu
 us base calls or 'N's in windowed segments. Sequencing: First-pass
 sequencing; ABI Prism 377 sequencer and analysis software.

updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 122-168, >(CAAA)n\$imple repeat
Seq primer: M13 Forward
POLYA=yes

FEATURES

Source

Location/Qualifiers
1..361
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UT-M-AM1-a-fv-h-01-0-UT"
/dev_stage="27-32 days"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: pRTD-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_MAM_N library is a normalized library constructed from mouse amygdala. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Waller Laboratories.
TAG_LIB=NIH_BMAP_MAM_N
TAG_TISSUE=amygdala
TAG_SEQ=GTGAG"

BASE COUNT 79 a 69 c 88 g 125 t
ORIGIN

Query Match 25.0%; Score 152; DB 9; Length 361;
Best Local Similarity 100.0%; Pred. No. 1.3e-68;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 429 TCGTCCGACCCCTACACACACTTCTCCAGCCGCCACTACGGGCTGTGTCAT 488
DB 358 TCGTCCGACCCCTACACACACTTCTCCAGCCGCCACTACGGGCTGTGTCAT 299
OY 489 GAATGCTTTGGCCCTTACCGATGCCAAGACTAGAGCCATGTGATGTGGTCT 548
DB 299 GAATGCTTTGGCCCTTACCGATGCCAAGACTAGAGCCATGTGATGTGGTCT 239
OY 549 CGTTCTTTCTTCAATGACGCTAGTATATACC 580
DB 238 CGTTCTTTCTTCAATGACGCTAGTATATACC 207

RESULT 6 425 bp mRNA linear EST 16-OCT-2001
LOCUS BB750925 RIKEN full-length enriched, pooled tissues, cerebellum,
DEFINITION etc. Mus musculus cDNA clone G130203B11 3', mRNA sequence.
ACCESSION BB750925
KEYWORDS BB750925.1 GI:16155161
SOURCE EST.

ORGANISM

Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 425)

REFERENCE
AUTHORS Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Inocenti, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Konda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watanishi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

JOURNAL
COMMENT Unpublished (2001)
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Ito, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujisake, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Kono, H., Fukushima, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based method for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES

Source

Location/Qualifiers
1..425
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G130203B11"
/clone_lib="RIKEN full-length enriched, pooled tissues, cerebellum, etc."
/note="pooled tissues: (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=male), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed)" (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)"
BASE COUNT 128 a 109 c 87 g 101 t
ORIGIN

Query Match 24.0%; Score 146; DB 10; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.9e-65;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 411 CAGTTACGACGAGTTATGCTGCGACCCCTACACACACACTTGTCCAGCCCCAC 470
DB 73 CAGTTACGACGAGTTATGCTGCGACCCCTACACACACACTTGTCCAGCCCCAC 132
OY 471 CTACGCGCTTGTGCTGCAATGCTTTGGCCCTTGAACGATGCAAGACTAGAGCA 530
DB 133 CTACGCGCTTGTGCTGCAATGCTTTGGCCCTTGAACGATGCAAGACTAGAGCA 192
OY 531 TCGTATGATGTGGTCTGCTCTTT 556
DB 193 TCGTATGATGTGGTCTGCTCTTT 218

RESULT 7

LOCUS

AA818804
DEFINITION AA818804/c 538 bp mRNA linear EST 03-JUL-1999
UI-R-A0-ar-b-11-0-UT.s1 UI-R-A0 Rattus norvegicus cDNA clone
a81c05.r1 Soares NHMpu SI Homo sapiens cDNA clone 1048424 5',
mRNA sequence.

ACCESSION AA818804
VERSION AA818804.1 GI:4227480
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

source

1. .430
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BG1-a1e-e-06-0-UI"
/clone_lib="NIH_BMAP_MSC_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MSC_N library is a normalized library constructed
from mouse spinal cord. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996. Tissue provided by Ms. Annie Novakovich,
Zivic-Willer Laboratories.
TAG_LIB=NIH_BMAP_MSC_N
TAG_TISSUE=spinal-cord
TAG_SEQ=TCGA"

BASE COUNT 97 a 85 c 104 g 144 t

ORIGIN

Query Match 27.9%; Score 170; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.8e-78;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 CAGTTAGGAGAGAGTTTNGCTGGCGAGCCCTACACACACACACTGTGCGAGCCCGAC 470
Db 370 CAGTTAGGAGAGAGTTTNGCTGGCGAGCCCTACACACACACACTGTGCGAGCCCGAC 311

QY 471 CTACGGCGCTTGCTGCATGAAATGCTTTTGCGCCCTTGACCGATGCCAAGACTAGAGGCA 530
Db 310 CTACGGCGCTTGCTGCATGAAATGCTTTTGCGCCCTTGACCGATGCCAAGACTAGAGGCA 251

QY 531 TGCCTGATATGTGGTCTCTCTCTTTCTTTCATTGACAGGCTAGATATACC 580
Db 250 TGCCTGATATGTGGTCTCTCTCTTTCTTTCATTGACAGGCTAGATATACC 201

RESULT 4
BF549922/c 533 bp mRNA linear EST 12-DEC-2000
LOCUS BF549922
DEFINITION UI-R-E0-bv-c-04-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
ACCESSION BF549922
VERSION BF549922.1 GI:11659652
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE 1 (bases 1 to 533)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: meoares@blue.weeg.uiowa.edu
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNLN (info@image.llnl.gov). IMAGE ID= 1777166
Seq primer: M13 forward.
Location/Qualifiers
1. .533

FEATURES
source

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

BASE COUNT 134 a 125 c 164 g 109 t 1 others

ORIGIN

Query Match 26.3%; Score 160; DB 12; Length 533;
Best Local Similarity 99.5%; Pred. No. 9.1e-73;
Matches 210; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 370 TACGCCAGCCGACCCCTGCGACTGCTGCTCTACAGTGACAGTTACGAGAGTTAT 429
Db 268 TACGCCAGCCGACCCCTGCGACTGCTGCTCTACAGTGACAGTTACGAGAGTTAT 209

QY 430 GCTGCCAGCCCTTACACACACACACTTGTCTCAGCCCCACCTACGGGCTTGTCATG 489
Db 208 GCTGCCAGCCCTTACACACACACACTTGTCTCAGCCCCACCTACGGGCTTGTCATG 149

QY 490 AATGCTTTGGCCCTTGACCGATGCCAAGACTAGAACCATGCTATGATGTGGCTTC 549
Db 148 AATGCTTTGGCCCTTGACCGATGCCAAGACTAGAACCATGCTATGATGTGGCTTC 89

QY 550 GTTCTTTCTTTCATTGACAGGCTAGATATACC 580
Db 88 GTTCTTTCTTTCATTGACAGGCTAGATATACC 58

RESULT 5
A1835011/c 361 bp mRNA linear EST 14-JUN-1999
LOCUS A1835011
DEFINITION UI-M-AM1-a1w-h-01-0-UI.s1 NIH_BMAP_MAM_N Mus musculus cDNA clone
ACCESSION A1835011
VERSION A1835011.1 GI:5469224
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 361)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized amygdala library cDNA library Preparation: M.B. Soares
Lab Clone distribution: NIH BMAP cDNA clones will be made available
by the means that is soon to be determined. When NIH determines the
means for distribution of the BMAP cDNA clones, this record will be

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 15, 2003, 15:59:00 ; Search time 856.787 Seconds

(without alignments)
11511.672 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactataataaaagcgcgc.....acaaccgttcgtccacat 609

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estlm:*
4: em_estlm:*
5: em_estlm:*
6: em_estlm:*
7: em_estlm:*
8: em_estlm:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estfun:*
17: gb_est8:*
18: em_gse_hum:*
19: em_gse_inv:*
20: em_gse_pln:*
21: em_gse_vrt:*
22: em_gse_fun:*
23: em_gse_mam:*
24: em_gse_mus:*
25: em_gse_other:*
26: em_gse_pro:*
27: em_gse_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	178	29.2	379	9 A1843387 UI-M-AQ1-
2	170	27.9	266	9 A1175239 EST218774
3	170	27.9	430	9 A1850237 UI-M-BG1-
4	160	26.3	533	12 BF549922 UI-R-B0-b
5	152	25.0	361	9 A1835011 UI-M-AM1-
6	146	24.0	425	10 BB750925 BB750925

C	7	92	15.1	538	9	AA818804	AA818804	UI-R-A0-a
8	92	15.1	600	13	BI988957	BI988957	4023-45 M	
9	80	13.1	408	12	BG795922	UTSW_SM38	BG795922	
10	70	11.5	228	14	Z28900	HSBB7F031 S	Z28900	
11	70	11.5	327	12	BG186689	RST5664 A	BG186689	
12	70	11.5	363	14	D60520	HUM14E06A	D60520	
13	70	11.5	367	9	AI095813	qB20911.x	AI095813	
14	70	11.5	363	9	AI189407	qB01h05.x	AI189407	
15	70	11.5	331	9	AI186273	qB20b08.x	AI186273	
16	70	11.5	392	9	AA975235	qB36C08.s	AA975235	
17	70	11.5	420	9	AI655094	wB67a10.x	AI655094	
18	70	11.5	427	12	BP223478	BP223478	7633e01.x	
19	70	11.5	439	9	AI968176	wu14a06.x	AI968176	
20	70	11.5	439	10	AMS89795	BG22d04.x	AMS89795	
21	70	11.5	439	10	BE501380	7a41h09.x	BE501380	
22	70	11.5	443	9	AA773715	af81c05.x	AA773715	
23	70	11.5	449	9	AI1374891	ca60912.x	AI1374891	
24	70	11.5	466	9	AI917298	te96c09.x	AI917298	
25	70	11.5	459	9	AI244212	qB86b11.x	AI244212	
26	70	11.5	477	10	AM197589	xm44g02.x	AM197589	
27	70	11.5	480	9	AI401040	ch27a12.x	AI401040	
28	70	11.5	495	9	AI656926	ct48e08.x	AI656926	
29	70	11.5	607	9	AU145000	AU145000	AU145000	
30	70	11.5	773	10	AV729198	AV729198	AV729198	
31	70	11.5	774	10	AV729057	AV729057	AV729057	
32	70	11.5	4039	11	BC026312	Homo_sapi	BC026312	
33	70	11.5	283	12	BG185609	RST4560 A	BG185609	
34	62	10.2	674	12	BG793800	UTSW_SM16	BG793800	
35	54	8.9	218	9	AI189257	gc99h03.x	AI189257	
36	54	8.9	251	14	F21037	HSPD0503 H	F21037	
37	54	8.9	372	9	AI799929	wc41a05.x	AI799929	
38	49	8.0	246	9	AI351484	q105e06.x	AI351484	
39	48	7.9	521	13	BM183349	fw10e03.x	BM183349	
40	48	7.9	723	12	BG306387	fm58d11.x	BG306387	
41	43	7.1	302	14	BM930547	UI-E-EJ1-	BM930547	
42	43	7.1	347	9	AI041587	ox56g05.x	AI041587	
43	43	7.1	428	9	AA460940	zx61d04.s	AA460940	
44	43	7.1	449	9	AI335996	qt42e01.x	AI335996	
45	43	7.1	451	9	AI193843	qet72h11.x	AI193843	

ALIGNMENTS

RESULT 1
A1843387/c 379 bp mRNA linear EST 14-JUL-1999
LOCUS A1843387
DEFINITION UI-M-AQ1-act-f-10-0-UI-s1 BMAP_MHI_N Mus musculus CDNA clone
UI-M-AQ1-act-f-10-0-UI 3', mRNA sequence.
ACCESSION A1843387
VERSION A1843387.1 GI:5477600
KEYWORDS EST.

ORGANISM

house mouse.

REFERENCE
Bonnafant, M.F., Lennon, G. and Soares, M.B.
1 (bases 1 to 379)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS
Bonnafant, M.F., Lennon, G. and Soares, M.B.

TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890

Email: mestr@nhi.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first A
strand cDNA and therefore this may represent a donalide poly A
tail. The sequence tag present in the CDNA between the NotI site

THIS PAGE BLANK (USPTO)

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12181; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95693 represent human amino acid sequences; and AAH13632 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 4684 BP; 1247 A; 1039 C; 1054 G; 1344 T; 0 other;
Query Match 3.3%; Score 20; DB 22; Length 4684;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 154 CTTGTATATACTTCTGCAAT 173
Db 3962 CTTGTATATACTTCTGCAAT 3981
RESULT 15
ABV23885
ID ABV23885 standard; cDNA; 4699 BP.
XX
AC ABV23885;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 23876.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-21314P.
PR 18-JUL-2000; 2000US-21907P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.

XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 4406-4407; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 4699 BP; 1250 A; 1040 C; 1057 G; 1347 T; 5 other;
Query Match 3.3%; Score 20; DB 23; Length 4699;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 154 CTTGTATATACTTCTGCAAT 173
Db 3977 CTTGTATATACTTCTGCAAT 3996
Search completed: March 15, 2003, 16:12:04
Job time : 157.485 secs

XX (HYSE-) HYSEQ INC.
PA
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue A, Yang Y, Wehrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR P-PSDB; AAM80096.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 4968; 6221pp; English.
XX
PS The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM79323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;
SQ
Query Match 7.1%; Score 43; DB 22; Length 1164;
Best Local Similarity 100.0%; Pred. No. 8.9e-12;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 100 CTGTTGTGCCAGGCCAACCCAGGAGATCTTCATGTACAGTG 142
Db 173 CTGTTGTGCCAGGCCAACCCAGGAGATCTTCATGTACAGTG 131
RESULT 13
AAS93634
ID AAS93634 standard; cDNA; 2118 BP.
XX
AC AAS93634;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29438.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
PI
DR WPI: 2001-639362/73.
DR P-PSDB; ABG29447.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 29438; 103pp; English.
XX
PS The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAK54197-AAK94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;
SQ
Query Match 6.4%; Score 39; DB 23; Length 2118;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 412 AGTTACGACGAGTTATGCGCGACCCCTACACACAC 450
Db 1300 AGTTACGACGAGTTATGCGCGACCCCTACACACAC 1338
RESULT 14
AAH14583
ID AAH14583 standard; cDNA; 4684 BP.
XX
AC AAH14583;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12181.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602

XX The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcripts. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition, to detect developmental specific genes, and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcc_sequences.
XX

Sequence 60 BP; 11 A; 14 C; 16 G; 19 T; 0 other;

Query Match 7.9%; Score 48; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.4e-14; Indels 0; Gaps 0;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGGTCTGTTCTTCT 558
Db 13 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGGTCTGTTCTTCT 60

RESULT 11
ABK34602
ID ABK34602 standard; cDNA; 3189 BP.

XX ABK34602;

XX 08-MAY-2002 (first entry)

DE Human cDNA for novel secreted protein, SEQ ID 371.

XX Human; 69; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune chryoditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.

XX Homo sapiens.

XX MO200177290-A2.

XX 18-OCT-2001.

XX 29-MAR-2001; 2001WO-US10295.

XX 06-APR-2000; 2000US-194941P.

XX (GEMV) GENETICS INST INC.

XX Wong GG, Clark HF, Fachtel K, Agostino MJ, Howes SH, Resnick RJ;

XX PI Gulukota K, Graham JR;

XX WPI; 2002-179323/23.

PT Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT -
XX
PS Claim 1; Page 188-189; 339pp; English.

XX The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins, their complements and sequences that hybridise to them.
CC Also included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the
CC polynucleotides, antibodies that bind to the proteins and identification
CC of modulators of the proteins or the expression of the polynucleotide.
CC The polynucleotides can be used as probes for the identification
CC and isolation of full length cDNA and genomic DNA. The polynucleotides
CC and proteins can also be used as nutritional supplements. The protein
CC is useful in the treatment of various immune deficiencies and disorders
CC such as viral infections, bacterial infections, fungal infections,
CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
CC autoimmune thyroiditis and diabetes) and allergic reactions and
CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.

Sequence 3189 BP; 1005 A; 554 C; 631 G; 999 T; 0 other;

Query Match 7.7%; Score 47; DB 24; Length 3189;
Best Local Similarity 100.0%; Pred. No. 8.1e-14; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 TGATGATGGGCTCGTTCTTTCTTTCATTCAGGCTAGTATATACC 580
Db 515 TGATGATGGGCTCGTTCTTTCTTTCATTCAGGCTAGTATATACC 561

RESULT 12
AAK53229/C
ID AAK53229 standard; cDNA; 1164 BP.

XX AAK53229;

XX 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 2758.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

XX MO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693325.

XX 30-NOV-2000; 2000US-0728422.

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
Query Match 11.5%; Score 70; DB 22; Length 1800;
Best Local Similarity 100.0%; Pred. No. 1.4e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 511 GATGCCAAGACTAGAGCCATGCTGATGCGTCTGTTCTTCTTCATTCGAGGCT 570
DB 1600 GATGCCAAGACTAGAGCCATGCTGATGCGTCTGTTCTTCTTCATTCGAGGCT 1659
QY 571 AGTATATACC 580
DB 1660 AGTATATACC 1669
RESULT 9
AAA07075
ID AAA07075 standard; cDNA; 2372 BP.
XX
AC AAA07075;
XX
DT 03-JUL-2000 (first entry)
XX
DE cDNA encoding human ataxin-2 binding protein (A2BP).
XX
KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
KW neuronal plasticity; cellular degeneration signal transduction pathway;
KW selective RNA transport; spinocerebellar ataxia type-2;
KW hyperproliferative disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 987..1979
FT /*tag= a
FT /product= "Human A2BP"
XX
PN WO200012710-A1.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20156.
XX
PR 01-SEP-1999; 98US-0145391.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Pulst SM, Shibata H;
XX
DR WPI: 2000-237873/20.
DR P-PSDB; AA81462.
XX
PT Nucleic acids encoding an ataxin-2 binding protein useful for
PT inhibiting the expression of active proteins from the SCA2 gene for the
PT treatment of spinocerebellar ataxia type-2 -
XX
PS Claim 6; Page 74-77; 82pp; English.
XX
CC This sequence represents cDNA encoding human ataxin-2 binding protein
CC (A2BP). Nucleotide sequences encoding human A2BP were originally
CC isolated in an adult brain cDNA library using the yeast two hybrid
CC method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
CC function that is encoded by the SCA2 gene located on chromosome 12. SCA2
CC has been linked to the autosomal dominant neurodegenerative disorder
CC spinocerebellar ataxia type-2. Individuals afflicted with the disease
CC exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
CC containing a polyglutamine stretch of about 35-39 residues, whereas that

CC of normal individuals contains approximately 22 contiguous glutamine
CC residues. A2BP and ataxin-2 are components of a cellular degeneration
CC signal transduction pathway. The pathogenic expanded form of ataxin-2 has
CC a higher affinity for A2BP relative to normal ataxin-2; the presence of
CC the expanded form is likely to promote degeneration. A2BP and ataxin have
CC also been found to have a role in gene regulation. The binding of A2BP to
CC ataxin-2 plays an important role in controlling gene expression via the
CC targeting of transport of specific RNAs, selective RNA transport being
CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
CC early in embryonic development. Both ataxin-2 and A2BP are able to bind
CC RNA, and are essential components of the RNA localisation network that
CC establishes cellular polarity in embryogenesis. In highly differentiated,
CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
CC function and are required for neuronal plasticity. A2BP nucleic acids may
CC be used for the recombinant production of A2BP proteins or fragments
CC thereof according to standard methodologies. For example, an A2BP protein
CC with an ataxin-2 or RNA binding capability but no signal transduction
CC function can be used as a dominant negative inhibitor of the cellular
CC degeneration signal transduction pathway. A2BP proteins with a signal
CC transduction function can be used to treat hyperproliferative disorders
CC (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX
SQ Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;
Query Match 11.5%; Score 70; DB 21; Length 2372;
Best Local Similarity 100.0%; Pred. No. 1.4e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 511 GATGCCAAGACTAGAGCCATGCTGATGCGTCTGTTCTTCTTCATTCGAGGCT 570
DB 2133 GATGCCAAGACTAGAGCCATGCTGATGCGTCTGTTCTTCTTCATTCGAGGCT 2192
QY 571 AGTATATACC 580
DB 2193 AGTATATACC 2202
RESULT 10
ABN34672
ID ABN34672 standard; DNA; 60 BP.
XX
AC ABN34672;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:7420.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-1B01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI: 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
PS Example 1; SEQ ID 7420; 47pp; English.

Query Match 11.5%; Score 70; DB 22; Length 607;
Best Local Similarity 100.0%; Pred. No. 1.3e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTCAGAGCT 570
DB 201 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTCAGAGCT 142

QY 571 AGTATATACC 580
DB 141 AGTATATACC 132

RESULT 7
AAH13824
ID AAH13824 standard; cDNA; 1513 BP.

AC AAH13824;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:10786.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saio K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX PS Claim 8; SEQ ID 10786; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNA easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AA992446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;

Query Match 11.5%; Score 70; DB 22; Length 1513;
Best Local Similarity 100.0%; Pred. No. 1.4e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTCAGAGCT 570
DB 1313 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTCAGAGCT 1312

QY 571 AGTATATACC 580
DB 1373 AGTATATACC 1382

RESULT 8
AAK52245
ID AAK52245 standard; cDNA; 1800 BP.

AC AAK52245;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 790.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX NO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Gao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AD, Yang Y, Wejrtman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX P-PsDB; AAM79112.
XX
XX PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX PS Claim 1; Page 2643-2645; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
XX encoded polypeptides (AAM78373-AAM80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or

ABK34530/c
 ID ABK34530 standard; cDNA; 539 BP.
 XX
 AC ABK34530;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human cDNA for novel secreted protein, SEQ ID 299.
 XX
 KM Human; ss; gene; secreted protein; immune deficiency; viral infection;
 KM bacterial infection; fungal infection; autoimmune disorder; burn;
 KM rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
 KM diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
 KM Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
 KM coagulation disorder; haemophilia; inflammatory disorder; ulcer;
 KM tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
 KM lymphoid cell deficiency.
 XX
 OS Homo sapiens.
 XX
 PN MO200177290-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 29-MAR-2001; 2001MO-US10295.
 XX
 PR 06-APR-2000; 2000US-194941P.
 XX
 PA (GENY) GENETICS INST INC.
 XX
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
 PI Gulkoka K, Graham JR;
 XX
 DR WPI: 2002-179323/23.
 XX
 PT Six hundred and twenty five polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders
 PT
 PS Claim 1; Page 153; 339pp; English.
 XX
 CC The invention relates to 625 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins, their complements and sequences that hybridise to them.
 CC Also included are a vector comprising the polynucleotide, a host cell
 CC transformed with the vector, the proteins encoded by the
 CC polynucleotides, antibodies that bind to the proteins and identification
 CC of modulators of the proteins or the expression of the polynucleotide.
 CC The polynucleotides can be used as probes for the identification
 CC and isolation of full length cDNA and genomic DNA. The polynucleotides
 CC and proteins can also be used as nutritional supplements. The protein
 CC is useful in the treatment of various immune deficiencies and disorders
 CC such as viral infections, bacterial infections, fungal infections,
 CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
 CC autoimmune thyroiditis and diabetes) and allergic reactions and
 CC conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment
 CC of burns, incisions and ulcers. The proteins are also useful for
 CC regulating haematopoiesis, for treating myeloid or lymphoid cell
 CC deficiencies. The present sequence is one of the 625 cDNA sequences
 CC encoding a secreted protein.
 XX
 SQ Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;
 Query Match 11.5%; Score 70; DB 24; Length 539;
 Best Local Similarity 100.0%; Pred. No. 1.3e-25;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 511 GATGCCAAGACTAGAGCCATGCTGATGTGGTCTCGTCTTCTTCATTGACAGCT 570

DB 191 GATGCCAAGACTAGAGCCATGCTGATGTGGTCTCGTCTTCTTCATTGACAGCT 132
 Oy 571 AGTATATACC 580
 DB 131 AGTATATACC 122
 RESULT 6
 AAH09205/c
 ID AAH09205 standard; cDNA; 607 BP.
 XX
 AC AAH09205;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer) SEQ ID NO:6040.
 XX
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KM Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 3; SEQ ID 6040; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 607 BP; 140 A; 152 C; 179 G; 129 T; 7 other;

XX 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #29436.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR P-PSDB; ABG29445.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 1; SEQ ID No 29436; 103bp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 406 BP; 94 A; 97 C; 85 G; 102 T; 28 other;
 SQ
 Query Match 11.5%; Score 70; DB 23; Length 406;
 Best Local Similarity 100.0%; Pred. No. 1.3e-25;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 511 GATGCCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTTCATTCGAGGCT 570
 DB 142 GATGCCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTTCATTCGAGGCT 201
 QY 571 AGTATATACC 580
 DB 202 AGTATATACC 211
 RESULT 4
 AAS93633 standard; cDNA; 481 BP.

XX AAS93633;
 AC 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #29437.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR P-PSDB; ABG29446.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 1; SEQ ID No 29437; 103bp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;
 SQ
 Query Match 11.5%; Score 70; DB 23; Length 481;
 Best Local Similarity 100.0%; Pred. No. 1.3e-25;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 511 GATGCCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTTCATTCGAGGCT 570
 DB 306 GATGCCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTTCATTCGAGGCT 365
 QY 571 AGTATATACC 580
 DB 366 AGTATATACC 375
 RESULT 5

PA (SCIO-) SCIOS INC.
 XX Stancon LW, White RT;
 XX WPI: 2002-010779/01.
 DR P-PSDB; AAU70146.
 XX Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases
 XX
 PS Claim 1: Fig 1: 189pp; English.
 XX
 CC The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 XX
 SQ Sequence 1340 BP, 366 A; 394 C; 317 G; 263 T; 0 other;
 Query Match 100.0%; Score 609; DB 24; Length 1340;
 Best Local Similarity 100.0%; Pred. No. 1.7e-30;
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGACTAATAAAAAGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTGG 60
 DB 535 ATGACTAATAAAAAGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTGG 594
 QY 61 GGGGGGGGTACAGCGCCGCACTTATGACAGGAGCGAGTGTGTGCCAGGCCAACCAG 120
 DB 595 GGGGGGGGTACAGCGCCGCACTTATGACAGGAGCGAGTGTGTGCCAGGCCAACCAG 654
 QY 121 GAGGATCTTCCATGATGAGTGCGCCAGTTCACTTGTATATATCTTCAATGCTTGGC 180
 DB 655 GAGGATCTTCCATGATGAGTGCGCCAGTTCACTTGTATATATCTTCAATGCTTGGC 714
 QY 181 TTTCCATATCCGGCCGCACTGTGTGACGTGATACCGAGGGGCTACCTTCAAGGCCGT 240
 DB 715 TTTCCATATCCGGCCGCACTGTGTGACGTGATACCGAGGGGCTACCTTCAAGGCCGT 774
 QY 241 GGTCCGACGGTGACAAACCTTCAGAGCTGGGGGCGCCCAACCCCAATCCCGGCTAT 300
 DB 775 GGTCCGACGGTGACAAACCTTCAGAGCTGGGGGCGCCCAACCCCAATCCCGGCTAT 834
 QY 301 GCGGAGTGTATCAAGAGCCAGTGTATGCAATTAATTTGTACAGGGTGTACGCT 360
 DB 835 GCGGAGTGTATCAAGAGCCAGTGTATGCAATTAATTTGTACAGGGTGTACGCT 894
 QY 361 GCATACCGCTACGCCAGCCACCCCTGCCACTGTGTGCTGCTTACAGTACAGTACGGA 420
 DB 895 GCATACCGCTACGCCAGCCACCCCTGCCACTGTGTGCTGCTTACAGTACAGTACGGA 954
 QY 421 CGAGTTATGTCGCGACCCCTACCAACACACTTGTCTCAGCCGCCACTTACGGCGTT 480
 DB 955 CGAGTTATGTCGCGACCCCTACCAACACACTTGTCTCAGCCGCCACTTACGGCGTT 1014
 QY 481 GGTGCCATGAATGCTTTTGGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATGAT 540
 DB 1015 GGTGCCATGAATGCTTTTGGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATGAT 1074
 QY 541 GTGGGTCTGTTTCTTCTTATGACAGGCTAGTATATACCAAGGGGATTCACCGTTT 600

DB 1075 GTGGGTCTGTTTCTTCTTATGACAGGCTAGTATATACCAAGGGGATACACCGTTT 1134
 QY 601 GGTCCATAT 609
 DB 1135 GGTCCATAT 1143
 RESULT 2
 ID AAK54001/c
 ID AAK54001 standard; cDNA; 327 BP.
 XX
 AC AAK54001;
 XX
 DT 16-NOV-2001 (first entry)
 XX
 DE Murine transcription associated protein encoding cDNA SEQ ID 566.
 XX
 KM Murine; liver; gene library; amino acid synthesis; binding protein;
 KM cell metabolism; energy metabolism; fatty acid metabolism; synthesis;
 KM phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;
 KM replication; transcription; translation; transport protein; ss.
 XX
 OS Mus musculus.
 XX
 FN DE20103510-UI.
 XX
 PD 07-JUN-2001.
 XX
 PF 28-FEB-2001; 2001DE-2003510.
 XX
 PR 02-DEC-1999; 99DE-1058160.
 XX
 PA (LION-) LION BIOSCIENCE AG.
 XX
 DR WPI: 2001-368570/39.
 XX
 PT Gene library containing sequences with specific 3'-ends and no polyA
 PT tail, encoding proteins involved in a wide range of cellular processes
 PT
 PS Claim 15; Page 174; 251pp; German.
 XX
 CC This invention describes a novel gene library (A) comprises a gene
 CC sequence (or its part) encoding a protein involved in amino acid
 CC synthesis, cellular/energy metabolism, metabolism of
 CC fatty acids/phospholipids, synthesis or breakdown of
 CC purines/pyrimidines/nucleosides/nucleotides, DNA
 CC replication/transcription/translation, or is a transport/binding protein.
 CC (A) are produced that correspond to the 3'-end of mRNA but without the
 CC polyA tail. They can be prepared more efficiently and with less effort
 CC than conventional libraries. AAK53436-AAK54275 represent fragments of the
 CC gene library described in the method of the invention.
 XX
 SQ Sequence 327 BP, 71 A; 65 C; 84 G; 107 T; 0 other;
 Query Match 14.1%; Score 86; DB 22; Length 327;
 Best Local Similarity 100.0%; Pred. No. 8.6e-34;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 495 TTTTGGCCCTTGACCGATGCCAAGACTAGGACCATGTATGATGAGGCTCTGTTCT 554
 DB 272 TTTTGGCCCTTGACCGATGCCAAGACTAGGACCATGTATGATGAGGCTCTGTTCT 213
 QY 555 TTCTTCATTGACGAGCTAGPATATACC 580
 DB 212 TTCTTCATTGACGAGCTAGPATATACC 187
 RESULT 3
 ID AAS93632
 ID AAS93632 standard; cDNA; 406 BP.
 XX
 AC AAS93632;

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:52:03 ; Search time 142.485 Seconds
(without alignments)
9625.317 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 1 atgactaataaaagccgt.....acaacgcttcctccatcat 609

Sequence: OLIGO NUC
Gapex 60.0 , Gapext 60.0

Scoring table: Gapex 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1 609 100.0 1340 24 AAS94693 Rat secreted facto
2 86 14.1 327 22 AAK54001 Murine transcripti
3 70 11.5 406 23 AAS93632 DNA encoding novel
4 70 11.5 481 23 AAS93633 DNA encoding novel
5 539 11.5 539 24 ABK34530 Human cDNA for nov
6 70 11.5 607 22 AAH09205 Human cDNA clone
7 70 11.5 1513 22 AAH13824 Human cDNA sequenc
8 70 11.5 1800 22 AAK52245 Human polynucleoti
9 2372 11.5 2372 21 AAA07075 cDNA encoding huma

10 48 7.9 60 24 ABN34672 Human spliced tran
11 47 7.7 3189 24 ABK34602 Human cDNA for nov
12 43 7.1 1164 22 AAK53229 Human polynucleoti
13 39 6.4 2118 23 AAS93634 DNA encoding novel
14 30 3.3 4684 22 AAH14583 Human cDNA sequenc
15 20 3.3 4699 23 ABV23385 Human prostate exp
16 20 3.3 4699 23 ABV23377 Human prostate exp
17 20 3.3 4699 23 ABV29766 Human prostate exp
18 20 3.3 4699 23 ABV29862 Human prostate exp
19 19 3.1 50 22 AAJ11184 Human SNP oligonuc
20 19 3.1 2547 23 ABH17965 Drosophila melanog
21 19 3.1 2564 23 ABH18561 Drosophila melanog
22 19 3.1 6846 23 ABH18560 Drosophila melanog
23 19 3.1 12631 23 ABH17964 Drosophila melanog
24 18 3.0 332 24 ABH19669 Prostate cancer re
25 18 3.0 372 23 AAS71730 DNA encoding novel
26 18 3.0 430 23 AAS66448 DNA encoding novel
27 18 3.0 755 23 AAS85509 DNA encoding novel
28 18 3.0 1668 23 AAS89718 DNA encoding novel
29 18 3.0 1668 23 AAS89818 DNA encoding novel
30 18 3.0 2262 23 ABH08285 Drosophila melanog
31 18 3.0 2830 24 AAS16455 Corn cDNA encoding
32 18 3.0 3015 23 AAS71731 DNA encoding novel
33 18 3.0 3431 22 AAS95513 First splice varia
34 18 3.0 3466 24 ABK6136 cDNA encoding huma
35 18 3.0 3468 22 AAS85514 Second splice vari
36 18 3.0 3582 22 AAS46157 Human DNA encoding
37 18 3.0 3741 23 ABH05203 Drosophila melanog
38 18 3.0 4130 23 ABH08284 Drosophila melanog
39 18 3.0 4830 20 ABA96894 Glucuronidase oxd
40 18 3.0 4830 21 AAC83156 DNA encoding Gluco
41 18 3.0 5920 23 ABH05202 Drosophila melanog
42 18 3.0 8159 22 AAJ19660 Human expressed po
43 18 3.0 8159 22 ABA06824 Human genomic DNA
44 18 3.0 8159 22 AAS92006 Genomic sequence #
45 18 3.0 8159 22 AAS9758 Human endocrine po

ALIGNMENTS

RESULT 1
AAS94693 standard; cDNA: 1340 BP.
ID AAS94693:
XX
AC AAS94693:
XX
DT 12-MAR-2002 (first entry)
XX
DE Rat secreted factor DNA clone P0184_D11 #1.
XX
KW Rat: secreted factor polypeptide; cardiac disease; kidney;
inflammatory disease; congestive heart failure; myocarditis; asthma; ss;
dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
renal infarction; hereditary nephritis; polycystic kidney disease;
chronic renal failure; renal vein thrombosis; medullary sponge kidney;
rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;
graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
Alzheimer's disease; gene therapy.
KW
XX
OS Rattus norvegicus.
XX
PN WO200174901-A2.
XX
PD 11-OCT-2001.
XX
PE 23-MAR-2001; 2001WO-US09555.
XX
PR 31-MAR-2000; 2000US-193548P.
PR 14-MAR-2001; 2001US-0809545.
XX

REFERENCE
AUTHORS
TITLE
JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115860)

Worley, K.C.

Direct Submission

Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Jul 18, 2002 this sequence version replaced gi:20514522.

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GWLC

Center clone name: CH230-34C2

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 99693 bases at least Q40

Consensus quality: 101936 bases at least Q20

Consensus quality: 102743 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 21 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1347: contig of 1347 bp in length

* 1348 1447: gap of unknown length

* 1448 2834: contig of 1387 bp in length

* 2835 2934: gap of unknown length

* 2935 4253: contig of 1319 bp in length

* 4254 4353: gap of unknown length

* 4354 4477: contig of 1124 bp in length

* 4478 5577: gap of unknown length

* 5578 7484: contig of 1907 bp in length

* 7485 7585: gap of unknown length

* 7586 9406: contig of 1822 bp in length

* 9407 9506: gap of unknown length

* 9507 11712: contig of 2206 bp in length

* 11713 11812: gap of unknown length

* 11813 13462: contig of 1650 bp in length

* 13463 13562: gap of unknown length

* 13563 17222: contig of 3660 bp in length

* 17223 17322: gap of unknown length

* 17323 20617: contig of 3295 bp in length

* 20618 20717: gap of unknown length

* 20718 24825: contig of 4108 bp in length

* 24826 24925: gap of unknown length

* 24926 28811: contig of 3886 bp in length

* 28812 28911: gap of unknown length

* 28912 33686: contig of 4775 bp in length

* 33687 33786: gap of unknown length

* 33787 40217: contig of 6431 bp in length

* 40218 40317: gap of unknown length

* 40318 47912: contig of 7595 bp in length

* 47913 48012: gap of unknown length

* 48013 56482: contig of 8470 bp in length

* 56483 56582: gap of unknown length

* 56583 66204: contig of 9622 bp in length

* 66205 66304: gap of unknown length

* 66305 74546: contig of 8242 bp in length

* 74547 74646: gap of unknown length

* 74647 86522: contig of 11876 bp in length

* 86523 86622: gap of unknown length

* 86623 100049: contig of 13427 bp in length
* 100050 100149: gap of unknown length
* 100150 115860: contig of 15711 bp in length.
Location/Qualifiers
1..115860
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-34C2"

FEATURES

source

BASE COUNT

33039 a 22417 c 22786 g 30574 t 7044 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 67; DB 2; Length 115860;

Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 AGGGGTGTTACGTCATACGCTACGCGCCAGCCACCCCTGCACTGCTGCTACA 406

DB 34697 AGGGGTGTTACGTCATACGCTACGCGCCAGCCACCCCTGCACTGCTGCTACA 34756

QY 407 GTGACAG 413

DB 34757 GTGACAG 34763

Search completed: March 15, 2003, 22:52:37

Job time : 1895.46 secs

repeat_region	complement(70389, .70704)	
repeat_region	/rpt_family="MER1"	
repeat_region	complement(72166, .72421)	
repeat_region	/rpt_family="MER33"	
repeat_region	72501, .72769	
repeat_region	/rpt_family="Alu"	
repeat_region	72843, .72956	
repeat_region	/rpt_family="MERS"	
repeat_region	complement(76303, .76564)	
repeat_region	/rpt_family="Alu"	
repeat_region	97534, .97700	
repeat_region	/rpt_family="MER3"	
repeat_region	complement(98469, .98526)	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(99089, .99381)	
repeat_region	/rpt_family="Alu"	
repeat_region	101648, .101930	
repeat_region	/rpt_family="Alu"	
repeat_region	102750, .103056	
repeat_region	/rpt_family="Alu"	
repeat_region	104010, .104277	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(111121, .111263)	
repeat_region	/rpt_family="MERS"	
repeat_region	complement(112263, .112476)	
repeat_region	/rpt_family="MER30"	
repeat_region	112556, .113234	
repeat_region	/rpt_family="Alu"	
repeat_region	114385, .114657	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(117031, .117194)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(118464, .118770)	
repeat_region	/rpt_family="Alu"	
repeat_region	119271, .119549	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(122831, .122940)	
repeat_region	/rpt_family="MER41"	
repeat_region	123471, .123742	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(129660, .129972)	
repeat_region	/rpt_family="Alu"	
repeat_region	135235, .135310	
repeat_region	/rpt_family="MER1"	
repeat_region	135874, .136157	
repeat_region	/rpt_family="Alu"	
repeat_region	139775, .140048	
repeat_region	/rpt_family="Alu"	
repeat_region	141734, .141814	
repeat_region	/rpt_family="MER"	
repeat_region	143071, .143352	
repeat_region	/rpt_family="Alu"	
repeat_region	148286, .148773	
repeat_region	/rpt_family="MER1"	
repeat_region	150431, .150707	
repeat_region	/rpt_family="Alu"	
repeat_region	155151, .155387	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(156101, .156213)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(156636, .156920)	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(159633, .159931)	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(160378, .160588)	
repeat_region	/rpt_family="Alu"	
repeat_region	161617, .162140	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(163200, .163315)	
repeat_region	/rpt_family="MER"	
BASE COUNT	48114 a 34256 c 33738 g	48430 t
ORIGIN		

Query Match	11.5%	Score 70	DB 9	Length 164538
Best Local Similarity	100.0%	Pred. No. 1.3e-29		
Matches 70	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy 511	GATGCACAGCTAGAGCAGCATGCTATGATGGCTCGTTCTTCTTCATTCAGAGCT	570		
Db 61328	GATGCCAAGACTAGAGCAGCATGCTATGATGGCTCGTTCTTCTTCATTCAGAGCT	61269		
Qy 571	AGTATATACC 580			
Db 61268	AGTATATACC 61259			
RESULT 15				
AC120661				
LOCUS				
DEFINITION	AC120661 115860 bp DNA linear HTG 23-JUL-2002			
ACCESSION	AC120661			
VERSION	AC120661.2 GI:21902877			
KEYWORDS	HTG; HTGS; PHASE1.			
SOURCE	Rattus norvegicus.			
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus			
REFERENCE	1 (bases 1 to 115860)			
AUTHORS	Munzy,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaralung,H.C., Are,J.R., Ayale,M., Banks,T., Barbataia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouch,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Buche,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davis,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Eatharth,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,N., Holloway,C., Hollins,B., Homesi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H., Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Matindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Mewner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogdu,N., Okunnu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sotakke,T., Spierke,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,F., Tameria,A., Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanai,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.			
TITLE	Unpublished			
JOURNAL	Direct Submission			
REFERENCE	2 (bases 1 to 115860)			
AUTHORS	Worley,K.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department			

BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN
Query Match 11.5%: Score 70; DB 9; Length 3343;
Best Local Similarity 100.0%; Pred. No. 1.4e-29;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 511 GATCCAGACTAGAGCCATGCTGATGAGTGGCTCGTTCTTTCTTCATTCGAGGCT 570
Db 1462 GATCCAGACTAGAGCCATGCTGATGAGTGGCTCGTTCTTTCTTCATTCGAGGCT 1521
Qy 571 AGTATATACC 580
Db 1522 AGTATATACC 1531
RESULT 14
AC005774/c 164538 bp DNA linear PRI 05-OCT-1998
LOCUS Homo sapiens chromosome 16, BAC clone 2603 (LANL), complete
DEFINITION
AC005774
VERSION AC005774.1 GI:3702264
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 164538)
Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Unpublished
2 (bases 1 to 164538)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Yeng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.
and Deaven,L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 164538)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Yeng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.
and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
SOURCE
1.164538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="2603"
repeat_region
1378..1665
/rpt_family="Alu"
repeat_region
complement(2144..2320)
/rpt_family="MER21"
repeat_region
complement(2198..2416)
/rpt_family="MER39"
repeat_region
complement(2450..2725)
/rpt_family="Alu"
repeat_region
complement(3157..3393)
/rpt_family="Alu"
repeat_region
complement(3524..3792)
/rpt_family="Alu"
5731..6020
/rpt_family="Alu"
repeat_region
complement(7557..7727)

/rpt_family="MER4"
8282..8928
/rpt_family="LTR8"
repeat_region
10055..10310
/rpt_family="THE1"
repeat_region
10472..10618
/rpt_family="L1"
repeat_region
10639..10939
/rpt_family="Alu"
repeat_region
complement(11109..11357)
/rpt_family="Alu"
12050..12312
/rpt_family="Alu"
repeat_region
complement(14642..14755)
/rpt_family="L1"
17429..18370
/rpt_family="MSTAR"
repeat_region
complement(19470..19819)
/rpt_family="MLT1"
repeat_region
complement(20352..20627)
/rpt_family="Alu"
23808..24058
/rpt_family="Alu"
repeat_region
27654..28217
/rpt_family="L1"
repeat_region
complement(32764..32956)
/rpt_family="L1"
repeat_region
complement(33392..33503)
/rpt_family="Alu"
34586..34670
/rpt_family="Alu"
repeat_region
34936..35489
/rpt_family="MLT2B2"
repeat_region
complement(36940..37172)
/rpt_family="Alu"
37173..37437
/rpt_family="THE1"
repeat_region
41428..41745
/rpt_family="THE1"
repeat_region
complement(42057..43927)
/rpt_family="L1"
43924..44555
/rpt_family="L1"
repeat_region
45128..45367
/rpt_family="Alu"
repeat_region
45467..45704
/rpt_family="Alu"
repeat_region
46192..46578
/rpt_family="MER25"
repeat_region
complement(46779..47141)
/rpt_family="THE1"
repeat_region
complement(48265..48550)
/rpt_family="Alu"
49630..50221
/rpt_family="MER41"
repeat_region
50781..51096
/rpt_family="Alu"
repeat_region
complement(51219..51439)
/rpt_family="Alu"
51462..51708
/rpt_family="MER33"
repeat_region
complement(52042..52321)
/rpt_family="Alu"
54001..54092
/rpt_family="Alu"
repeat_region
complement(55760..56014)
/rpt_family="Alu"
56616..56899
/rpt_family="Alu"
repeat_region
complement(66463..66747)
/rpt_family="Alu"
repeat_region
complement(66908..67180)
/rpt_family="Alu"

REFERENCE 2 (pages 1 to 2279)
AUTHORS Shibata, H., Huynh, D. P., Vo, T. T. and Puls, S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics

Submitted (16-NOV-1998) Division of Neurology and Neurogenetics Laboratory, Burns and Allen Research Institute, Cedars-Sinai Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA

FEATURES	Location/Qualifiers
source	1. .2279

```

gene      /note="similar to Homo sapiens cDNA clone HSP38A20"
1.2279
/feature="A2Bp"
1047. 2180
CDS

```

```
misc_feature
/crm:slat:ion="MAQOPYASAOXFPAPNDGAPBAETATAPHPHAEPEYGOCTVPEHTLANT
LYPPATQTSBESPADTSAQTYUGTATPDNDGAPDGOPTOQESIENTEKQSOPKRLTHAS
NIPFFRBDPLDKOMFGQKILDBAIIITENEGSKQFEGVTTENSADADRAREKHAHTGVY
VEGRKILEVMNTAPRWTKTKPTATPYTNWKNMLPVGAAYSPFPAAGVLLAQNOABESQ
SMYASPSLVTYSANRPPYPATPAATTAAYRGLHIEGRGRTYVNTFRVALLCPQNP
GVVYODGFGADIVGGVAAYRAOCTPPTAAIASYOSYGRVVAADPYHHALAPATYGVG
GAMNA.FALULTDIAKTSHADVUGLVLTSLQASLYRCGVNRFAPY"
1341..1358
```

```
misc_feature      /note="Region: RNA binding motif RNP-2"
1452. 1472
```

```

/Note="Region: RNA binding motif RNP-1
BASE COUNT      497 a      712 c      636 g      434 t
ORIGIN

```

Query Match	11.5%	Score 70;	DB 9;	Length 2279;
Best Local Similarity	100.0%	Pred. No. 1.5e-29;		
Matches	70;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Oy 511 GATGCCAAGCTAGGACCCATGCTGATGATGTGGGTCTCGTTCTTTCATTTCAGGCT 570
Db 2079 GATGCCAAGCTAGGACCCATGCTGATGATGTGGGTCTCGTTCTTTCATTTCAGGCT 2138

QY	571	AGTATATACC	580
Db	2139	AGTATATACC	2148

RESULT 11	
ARI34676	
LOCUS	2372 bp
ARI34676	DNA
	linear
	PAT 16-MAY-2001

ACCESSION	AR134676
VERSION	AR134676.1
	GI:14123581

SOURCE ORGANISM	Unknown.
	Unknown.

REFERENCE	1 (bases 1 to 2372)
AUTHORS	Pulst, S.M. and Shibata, H

JOURNAL Patent: US 6194171-A 1 27-FEB-2001;
FEATURES Location/Qualifiers

```

/organism="unknown"
BASE COUNT      548 a      726 c      646 g
ORIGIN

```

```
Query Match      11.5%; Score 70; DB 6; Length 2372;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

Accession	Sequence	Position
OY	GATCCAAAGCTGAGAGCCATGCTGATGATGAGGGCTCGTCTTTCTTCATGAGGCT	570
OY	511	
Db	GATCCAAAGCTGAGAGCCATGCTGATGATGAGGGCTCGTCTTTCTTCATGAGGCT	2192
Db	2133	
OY	AGTATATACC	580
OY	571	
Db	AGTATATACC	2202
Db	2193	

RESULT	12
HSFOX14	
LOCUS	2971 bp
DEFINITION	DNA linear PRI 01-MAR-2002
	HsFOX14
	Homo sapiens hexaribonucleotide binding protein 1 alpha, beta, and gamma

gamma isoforms (HKNBP1) gene, alternatively spliced, exon 14 and partial cds and complete cds.
 AF109120
 AF109120.1 GI:19032383

SEGMENT	14 of 14
SOURCE	Homo sapiens

ORGANISM	Homosapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (1994-1997)		

AUTHORS Chen, W. and Winkelmann, J. C.
TITLE The Exon-Intron Organization of the Human HRNP1 Gene

REFERENCE
AUTHORS
2 (bases 1 to 2971)
Chen, W., Chu, Z.-L., Blough, R. I., Liu, L., Hoppes, B. and
Winkelmann, T. C.

TITLE
Molecular Cloning and Chromosomal Localization of a Novel Human Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene Homologous to fox-1 in *Caenorhabditis elegans*

REFERENCE	unpublished
3 (bases 1 to 2971)	
Chen W and Winkelmann J.C.	

TITLE Internal Medicine/Hematology-Oncology,
JOURNAL Submitted (25-NOV-1998)
DIRECT SUBMISSION University of Cincinnati College of Medicine, 231 Bethesda Ave. ML

FEATURES	Location/Qualifiers
0508, Cincinnati, OH 45207-0508, OH	
1-2971	

gene

mRNA

mRNA

mRNA

/Gene: hNF1A
 /Product: hexaribonucleotide binding protein 1 isoform gamma"
 join(AFI09107.1:372..711,AFI09108.1:173..415,
 AFI09109.1:166..309,AFI09110.1:94..147,AFI09111.1:75..167,
 AFI09112.1:50..119,AFI09113.1:200..253,AFI09114.1:32..102,
 AFI09115.1:67..139,AFI09116.1:108..150,AFI09117.1:35..99,
 AFI09118.1:67..119,AFI09119.1:108..150,AFI09120.1:35..99,
 AFI09121.1:67..119,AFI09122.1:108..150,AFI09123.1:35..99,
 AFI09124.1:67..119,AFI09125.1:108..150,AFI09126.1:35..99,
 AFI09127.1:67..119,AFI09128.1:108..150,AFI09129.1:35..99,
 AFI09130.1:67..119,AFI09131.1:108..150,AFI09132.1:35..99,
 AFI09133.1:67..119,AFI09134.1:108..150,AFI09135.1:35..99,
 AFI09136.1:67..119,AFI09137.1:108..150,AFI09138.1:35..99,
 AFI09139.1:67..119,AFI09140.1:108..150,AFI09141.1:35..99,
 AFI09142.1:67..119,AFI09143.1:108..150,AFI09144.1:35..99,
 AFI09145.1:67..119,AFI09146.1:108..150,AFI09147.1:35..99,
 AFI09148.1:67..119,AFI09149.1:108..150,AFI09150.1:35..99,
 AFI09151.1:67..119,AFI09152.1:108..150,AFI09153.1:35..99,
 AFI09154.1:67..119,AFI09155.1:108..150,AFI09156.1:35..99,
 AFI09157.1:67..119,AFI09158.1:108..150,AFI09159.1:35..99,
 AFI09160.1:67..119,AFI09161.1:108..150,AFI09162.1:35..99,
 AFI09163.1:67..119,AFI09164.1:108..150,AFI09165.1:35..99,
 AFI09166.1:67..119,AFI09167.1:108..150,AFI09168.1:35..99,
 AFI09169.1:67..119,AFI09170.1:108..150,AFI09171.1:35..99,
 AFI09172.1:67..119,AFI09173.1:108..150,AFI09174.1:35..99,
 AFI09175.1:67..119,AFI09176.1:108..150,AFI09177.1:35..99,
 AFI09178.1:67..119,AFI09179.1:108..150,AFI09180.1:35..99,
 AFI09181.1:67..119,AFI09182.1:108..150,AFI09183.1:35..99,
 AFI09184.1:67..119,AFI09185.1:108..150,AFI09186.1:35..99,
 AFI09187.1:67..119,AFI09188.1:108..150,AFI09189.1:35..99,
 AFI09190.1:67..119,AFI09191.1:108..150,AFI09192.1:35..99,
 AFI09193.1:67..119,AFI09194.1:108..150,AFI09195.1:35..99,
 AFI09196.1:67..119,AFI09197.1:108..150,AFI09198.1:35..99,
 AFI09199.1:67..119,AFI09200.1:108..150,AFI09201.1:35..99,
 AFI09202.1:67..119,AFI09203.1:108..150,AFI09204.1:35..99,
 AFI09205.1:67..119,AFI09206.1:108..150,AFI09207.1:35..99,
 AFI09208.1:67..119,AFI09209.1:108..150,AFI09210.1:35..99,
 AFI09211.1:67..119,AFI09212.1:108..150,AFI09213.1:35..99,
 AFI09214.1:67..119,AFI09215.1:108..150,AFI09216.1:35..99,
 AFI09217.1:67..119,AFI09218.1:108..150,AFI09219.1:35..99,
 AFI09220.1:67..119,AFI09221.1:108..150,AFI09222.1:35..99,
 AFI09223.1:67..119,AFI09224.1:108..150,AFI09225.1:35..99,
 AFI09226.1:67..119,AFI09227.1:108..150,AFI09228.1:35..99,
 AFI09229.1:67..119,AFI09230.1:108..150,AFI09231.1:35..99,
 AFI09232.1:67..119,AFI09233.1:108..150,AFI09234.1:35..99,
 AFI09235.1:67..119,AFI09236.1:108..150,AFI09237.1:35..99,
 AFI09238.1:67..119,AFI09239.1:108..150,AFI09240.1:35..99,
 AFI09241.1:67..119,AFI09242.1:108..150,AFI09243.1:35..99,
 AFI09244.1:67..119,AFI09245.1:108..150,AFI09246.1:35..99,
 AFI09247.1:67..119,AFI09248.1:108..150,AFI09249.1:35..99,
 AFI09250.1:67..119,AFI09251.1:108..150,AFI09252.1:35..99,
 AFI09253.1:67..119,AFI09254.1:108..150,AFI09255.1:35..99,
 AFI09256.1:67..119,AFI09257.1:108..150,AFI09258.1:35..99,
 AFI09259.1:67..119,AFI09260.1:108..150,AFI09261.1:35..99,
 AFI09262.1:67..119,AFI09263.1:108..150,AFI09264.1:35..99,
 AFI09265.1:67..119,AFI09266.1:108..150,AFI09267.1:35..99,
 AFI09268.1:67..119,AFI09269.1:108..150,AFI09270.1:35..99,
 AFI09271.1:67..119,AFI09272.1:108..150,AFI09273.1:35..99,
 AFI09274.1:67..119,AFI09275.1:108..150,AFI09276.1:35..99,
 AFI09277.1:67..119,AFI09278.1:108..150,AFI09279.1:35..99,
 AFI09280.1:67..119,AFI09281.1:108..150,AFI09282.1:35..99,
 AFI09283.1:67..119,AFI09284.1:108..150,AFI09285.1:35..99,
 AFI09286.1:67..119,AFI09287.1:108..150,AFI09288.1:35..99,
 AFI09289.1:67..119,AFI09290.1:108..150,AFI09291.1:35..99,
 AFI09292.1:67..119,AFI09293.1:108..150,AFI09294.1:35..99,
 AFI09295.1:67..119,AFI09296.1:108..150,AFI09297.1:35..99,
 AFI09298.1:67..119,AFI09299.1:108..150,AFI09300.1:35..99,
 AFI09301.1:67..119,AFI09302.1:108..150,AFI09303.1:35..99,
 AFI09304.1:67..119,AFI09305.1:108..150,AFI09306.1:35..99,
 AFI09307.1:67..119,AFI09308.1:108..150,AFI09309.1:35..99,
 AFI09310.1:67..119,AFI09311.1:108..150,AFI09312.1:35..99,
 AFI09313.1:67..119,AFI09314.1:108..150,AFI09315.1:35..99,
 AFI09316.1:67..119,AFI09317.1:108..150,AFI09318.1:35..99,
 AFI09319.1:67..119,AFI09320.1:108..150,AFI09321.1:35..99,
 AFI09322.1:67..119,AFI09323.1:108..150,AFI09324.1:35..99,
 AFI09325.1:67..119,AFI09326.1:108..150,AFI09327.1:35..99,
 AFI

LOCUS AK001027 1513 bp mRNA linear PRI 01-AUG-2002
 DEFINITION Homo sapiens cDNA FLJ10165 fis, clone HEMBA1003591, weakly similar
 to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.
 AK001027
 ACCESSION AK001027.1 GI:7022045
 VERSION AK001027.1
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
 mRNA, clone_11b:HEMBA1 clone:HEMBA1003591.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Iisogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
 Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
 Magatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
 Sugawara, M., Takahashi, M., Chiba, Y., Ichida, S., Murakami, K.,
 Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
 Ichii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
 Nakamura, Y., Nagahara, K., Masuno, Y., Ninomiya, K., and Iwayanagi, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 1513)
 Iisogai, T. and Otsuki, T.
 Direct Submission
 Submitted (16-FEB-2000) Takao Iisogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-338-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan: cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5' - 3' end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.
 FEATURES
 source
 location/Qualifiers
 1.1513
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1003591"
 /cissue_type="whole embryo, mainly head"
 /clone_id="HEMBA1"
 /dev_stage="embryo, 10 weeks"
 /note="cloning vector: pME18SFL3"
 221.1414
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91472.1"
 /db_xref="GI:7022046"
 /translation="MNCERQLRGNQEAAPPTMAQPVASQFAPPONGIPAEVTAP
 HPHAPBEYTGQTVPEHTLNLVPPAQTHSEOSPADTSAQVSGTADADDAAPTDGQ
 QTPSEVENTENKQPKRLHVNIPRRFPDLROMFGQIGLIVEIIFNRSKGFQF
 VTFENSADTARERKELHGVTEGRKTEVNNAATKVTNAPYNGMLNIVNGV
 XSPREYAGTVLLCOANOEGSSMSAPSLVYTGAMGEPYPATATAAAYGALRGRG
 RTVNTFRAAPPPIPAVGAVYDDFGVYDGFYGGYARIVAOPTPATTAASDSIG
 RYVAADPYHHLAPAPYGVGANNAPFLDTAKTRSHLDVGLVLSLQASIVRGYN
 RFAPY"
 CDS
 BASE COUNT 354 a 418 c 395 g 346 t
 ORIGIN
 Query Match 11.5%; Score 70; DB 9; Length 1513;
 Best Local Similarity 100.0%; Pred. No. 1.5e-29;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 511 GATGCCAAGACTAGAGCCATGCTGATGATGGGTCCTCTTTCTTTCATTGCAGGCT 570
 Db 1313 GATGCCAAGACTAGAGCCATGCTGATGATGGGTCCTCTTTCTTTCATTGCAGGCT 1372
 Oy 571 AGTATATACC 580
 Db 1373 AGTATATACC 1382
 RESULT 9

AF229057
 LOCUS AF229057 2000 bp mRNA linear PRI 01-FEB-2002
 DEFINITION Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
 (HRNBP1) mRNA, complete cds.
 AF229057
 ACCESSION AF229057
 VERSION AF229057.1 GI:18461366
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Chen, W. and Winkelmann, J. C.
 Direct Submission
 Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
 University of Cincinnati College of Medicine, The Vontz Center for
 Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
 FEATURES
 source
 location/Qualifiers
 1.2000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16p13.3"
 1.2000
 /gene="HRNBP1"
 256.1512
 /gene="HRNBP1"
 /note="RNA-binding protein"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform
 gamma"
 /protein_id="AAU71904.1"
 /db_xref="GI:18461367"
 /translation="MLASQGVLLHPYGVEMIVPAAPYLPGLIQGNQEAAPPTMAQ
 VYASQFAPPONGIPAEVTAPHPHAPBEYTGQTVPEHTLNLVPPAQTHSEOSPADTSA
 QVSGTATDTPDPAFTDQGPOTQSPSTENKQPKRLHVNIPRRFPDLROMFGQF
 GILIDVEIIFNRSKGFQFVTFENSADTARERKELHGVTEGRKTEVNNAATKVTN
 KTVNAPYNGMLNIVNGVXSPREYAGTVLLCOANOEGSSMSAPSLVYTGAMGEP
 YPATATAAAYGALRGRGRTVNTFRAAPPPIPAVGAVYDDFGVYDGFYGGYARIVAOPT
 PATTAASDSIGRYVAADPYHHLAPAPYGVGANNAPFLDTAKTRSHLDVGLVLSLQASIVRGYN
 RFAPY"
 CDS
 BASE COUNT 502 a 505 c 487 g 506 t
 ORIGIN
 Query Match 11.5%; Score 70; DB 9; Length 2000;
 Best Local Similarity 100.0%; Pred. No. 1.5e-29;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 511 GATGCCAAGACTAGAGCCATGCTGATGATGGGTCCTCTTTCTTTCATTGCAGGCT 570
 Db 1411 GATGCCAAGACTAGAGCCATGCTGATGATGGGTCCTCTTTCTTTCATTGCAGGCT 1470
 Oy 571 AGTATATACC 580
 Db 1471 AGTATATACC 1480
 RESULT 10
 AF107203
 LOCUS AF107203 2279 bp mRNA linear PRI 23-JUN-2000
 DEFINITION Homo sapiens ataxin-2-binding protein (A2BP) mRNA, complete cds.
 AF107203
 ACCESSION AF107203.1 GI:8671585
 VERSION AF107203.1
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Shibata, H., Huynh, D. P., Vo, T. T., and Puls, S.-M.
 Title A novel protein, A2BP, with RNA binding motif binds to C-terminal
 ataxin-2
 Unpublished

* 32686 34092: contig of 1407 bp in length
* 34093 34192: gap of unknown length
* 34193 34991: contig of 1299 bp in length
* 35492 35991: gap of unknown length
* 35992 37096: contig of 1505 bp in length
* 37097 37196: gap of unknown length
* 37197 38312: contig of 1116 bp in length
* 38313 38412: gap of unknown length
* 38413 40579: contig of 2167 bp in length
* 40580 40679: gap of unknown length
* 40680 42711: contig of 2032 bp in length
* 42712 42811: gap of unknown length
* 42812 44224: contig of 1413 bp in length
* 44225 44324: gap of unknown length
* 44325 45859: contig of 1535 bp in length
* 45860 45959: gap of unknown length
* 45960 47975: contig of 2016 bp in length
* 47976 48075: gap of unknown length
* 48076 50049: contig of 1974 bp in length
* 50050 50149: gap of unknown length
* 50150 51700: contig of 1551 bp in length
* 51701 51800: gap of unknown length
* 51801 53551: contig of 1751 bp in length
* 53552 53651: gap of unknown length
* 53652 56127: contig of 2476 bp in length
* 56128 56227: gap of unknown length
* 56228 57920: contig of 1693 bp in length
* 57921 58020: gap of unknown length
* 58021 60257: contig of 2237 bp in length
* 60258 60357: gap of unknown length
* 60358 62838: contig of 2481 bp in length
* 62839 62938: gap of unknown length
* 62939 64839: contig of 1901 bp in length
* 64840 64939: gap of unknown length
* 64940 66877: contig of 1938 bp in length
* 66878 66977: gap of unknown length
* 66978 69856: contig of 2879 bp in length
* 69857 71716: gap of unknown length
* 71717 71816: contig of 1760 bp in length
* 71817 73929: gap of unknown length
* 73929: contig of 2113 bp in length.
Location/Qualifiers
1..73929
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-33J14"
BASE COUNT 19038 a 16649 c 14558 g 19363 t 4321 others
ORIGIN
Query Match 13.0% Score 79; DB 2; Length 73923;
Best Local Similarity 100.0%; Pred. No. 5.9e-35;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 411 CAGTTACGACGAGTTATGTCGCCGACCCCTTACACACACACACTGCTCCAGCCCCCAG 470
DB 11590 CAGTTACGACGAGTTATGTCGCCGACCCCTTACACACACACACTGCTCCAGCCCCCAG 11649
QY 471 CTACGCGCTTGTCGCATG 489
DB 11650 CTACGCGCTTGTCGCATG 11668
RESULT 7
AB060859 1475 bp mRNA linear PRI 13-JUN-2001
LOCUS AB060859 Macaca fascicularis brain cDNA clone:Qtra-11594, full insert
DEFINITION Macaca fascicularis brain cDNA clone:Qtra-11594, full insert
sequence.
ACCESSION AB060859
VERSION AB060859.1 GI:13874510
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Macaca fascicularis adult male temporal lobe right cDNA to mRNA,
clone_lib:macaque brain cDNA library Qtra clone:Qtra-11594.
ORGANISM Macaca fascicularis

REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.
AUTHORS Oada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
TITLE Isolation of full-length cDNA clones from macaque brain cDNA
libraries
JOURNAL Unpublished
AUTHORS Hashimoto, K., Oada, N., Hida, M., Kusuda, J. and Sugano, S.
TITLE Direct Submission (27-APR-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinyuku-ku, Tokyo 162-8640, Japan
(E-mail: hashim@nih.go.jp, URL: http://www.nih.go.jp/yoken/genbank/
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
COMMENT Lab host: TOP10
Vector: pME185-FL3 (Acc. No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CAGCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGTCGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME185-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by oligo-capping method
(Sugano et al., Institute of Medical Science, University of
Tokyo).
Custom primer used for sequencing
(5' end primer [CTTCTGCTTAAAGCTGCG] ;
3' end primer [CGACGACGCTGACACAC]).
Location/Qualifiers
1..1475
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="Qtra-11594"
/sex="male"
/tissue-type="temporal lobe right"
/clone_lib="macaque brain cDNA library Qtra"
/dev_stage="adult"
90..1220
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB46877.1"
/db_xref="GI:13874511"
/translation="MEERGRKRVQOQNOEAAADPTMAQPYASNOFAPRPGICPAEYT
APHPAPAEYGTTPVEHTLNLPAPQTHSEGSADPTNAGTVSGTATQDDAPTDG
QPTQPSBENTKSKQPRKLHVSNIIPFRFDPDLROMQGFQKILDELIIENERSKGF
GFVFENSADADRAREKLHGVGRKLVESGRKLVNARVNTKNLAPYTNKWLNVVVG
AVGSEPFYAGTVLILCOANOEGSSVMGAPSSLVNTSAMGPFYPATATAAAYRGAHLRG
RGRTVNTFRRAAPPPPIPRATGVTYODGDFGADITGCVAYRAOPTPRAPAAVSDR
NOFVFVADEISCTNSAVTDFEFLPPTTHLLOPPPALVP"
BASE COUNT 389 a 414 c 363 g 309 t
ORIGIN
Query Match 11.5% Score 70; DB 9; Length 1475;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 511 GATCCCAAGACTAGAGCATGCTGATGATGAGCTCTGTTCTTTTCATTCAGAGCT 570
DB 1241 GATCCCAAGACTAGAGCATGCTGATGATGAGCTCTGTTCTTTTCATTCAGAGCT 1300
QY 571 AGTATATACC 580
DB 1301 AGTATATACC 1310
RESULT 8
AK001027

QY 549 CGTCTTCTTCATTCGAGCTAGTATATACC 580
 Db 112495 CGTCTTCTTCATTCGAGCTAGTATATACC 112464
 RESULT 6
 AC103207
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-33J14, *** SEQUENCING IN PROGRESS
 ***, 44 unordered pieces.
 AC103207
 VERSION
 AC103207.3 GI:21731168
 KEYWORDS
 HTG; HTGS_PHASE1.
 SOURCE
 Norway rat.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 73929)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Albreck, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T.,
 Barbieri, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Bukey, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J.S., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Duh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Edmar, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falla, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabioli, A., Gao, J., Garcia, A., Garner, T., Garzon, G., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loueseg, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawlin, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokwenkwo, S., Oguh, M., Okunolu, G.,
 Otagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, K.,
 Sutton, A., Svatek, A., Tabor, P., Tameria, A., Tameria, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinton, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczek, R., Woode, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 73929)
 Worley, K.C.
 Direct Submission
 Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 73929)
 Worley, K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 11, 2002 this sequence version replaced gi:17974697.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GJUL
 Center clone name: CH230-33J14
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 23573 bases at least Q40
 Consensus quality: 25350 bases at least Q30
 Consensus quality: 26823 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length.
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 44 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 1065: contig of 1065 bp in length
 1066 1165: gap of unknown length
 1166 2411: contig of 1246 bp in length
 2412 2511: gap of unknown length
 2512 3962: contig of 1471 bp in length
 3963 4082: gap of unknown length
 4083 5596: contig of 1514 bp in length
 5597 5696: gap of unknown length
 5697 7081: contig of 1385 bp in length
 7082 7181: gap of unknown length
 7182 8497: contig of 1315 bp in length
 8498 8596: gap of unknown length
 8597 10100: contig of 1504 bp in length
 10101 10200: gap of unknown length
 10201 11249: contig of 1049 bp in length
 11250 11350: gap of unknown length
 11351 12701: contig of 1352 bp in length
 12702 12801: gap of unknown length
 12802 14312: contig of 1411 bp in length
 14313 14313: gap of unknown length
 14314 15380: contig of 1068 bp in length
 15381 15480: gap of unknown length
 15481 16516: contig of 1036 bp in length
 16517 17946: gap of unknown length
 17947 18046: contig of 1330 bp in length
 18047 19146: gap of unknown length
 19147 19246: contig of 1100 bp in length
 19247 20526: gap of unknown length
 20527 21730: contig of 1280 bp in length
 21731 21731: gap of unknown length
 21732 21730: contig of 1104 bp in length
 21731 21830: gap of unknown length
 21831 23358: gap of 1528 bp in length
 23359 23458: gap of unknown length
 23459 25338: contig of 1880 bp in length
 25339 25438: gap of unknown length
 25439 26638: contig of 1200 bp in length
 26639 26738: gap of unknown length
 26739 28368: contig of 1630 bp in length
 28369 28468: gap of unknown length
 28469 29896: contig of 1428 bp in length
 29897 31050: gap of unknown length
 31051 31050: contig of 1054 bp in length
 31051 31151: gap of unknown length
 31151 32585: contig of 1435 bp in length
 32586 32685: gap of unknown length

VERSION AC120661.2 GI:21902877
KEYWORDS HTG: HTGS PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 115860)
AUTHORS Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayete,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeValla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotco,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homel,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudah,S.,
Karlsen,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Melker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moer,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,
Orgunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sison,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,
Uemami,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
TITLE Unpublished
JOURNAL 2 (bases 1 to 115860)
REFERENCE Direct Submission
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115860)
REFERENCE Direct Submission
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20514522.
COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXIC
Center clone name: CH230-34C2
----- Summary Statistics
Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye 100# of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 99693 bases at least Q40
Consensus quality: 101936 bases at least Q30
Consensus quality: 102743 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1347: contig of 1347 bp in length
1348 1447: gap of unknown length
1349 1448: contig of 1387 bp in length
1350 2835 2934: gap of unknown length
1351 2935 4253: contig of 1319 bp in length
1352 4254 4354: gap of unknown length
1353 4355 5477: contig of 1124 bp in length
1354 5478 5577: gap of unknown length
1355 5578 7484: contig of 1907 bp in length
1356 7485 7585: gap of unknown length
1357 7586 9406: contig of 1822 bp in length
1358 9407 9506: gap of unknown length
1359 9507 11712: contig of 2206 bp in length
1360 11713 11812: gap of unknown length
1361 11813 14662: contig of 1650 bp in length
1362 14663 13562: gap of unknown length
1363 13563 17222: contig of 3660 bp in length
1364 17223 17322: gap of unknown length
1365 17323 20617: contig of 3295 bp in length
1366 20618 20717: gap of unknown length
1367 20718 24825: contig of 4108 bp in length
1368 24826 24925: gap of unknown length
1369 24926 28811: contig of 3886 bp in length
1370 28812 28911: gap of unknown length
1371 28912 33686: contig of 4775 bp in length
1372 33687 33786: gap of unknown length
1373 33787 40217: contig of 6431 bp in length
1374 40218 40318: gap of unknown length
1375 40319 47912: contig of 7595 bp in length
1376 47913 48012: gap of unknown length
1377 48013 56482: contig of 8470 bp in length
1378 56483 56582: gap of unknown length
1379 56583 66204: contig of 9622 bp in length
1380 66205 66304: gap of unknown length
1381 66305 74546: contig of 8242 bp in length
1382 74547 74646: gap of unknown length
1383 74647 86522: contig of 11876 bp in length
1384 86523 86622: gap of unknown length
1385 86623 100049: contig of 13427 bp in length
1386 100050 100149: gap of unknown length
1387 100150 115860: contig of 15711 bp in length.

FEATURES

source location/Qualifiers
1..115860
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-34C2"

BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others
ORIGIN

Query Match 15.1%; Score 92; DB 2; Length 115860;
Best Local Similarity 100.0%; Pred. No. 1.1e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

489 GAATGTTTGGCCCTTACCGATGCGACGACTGAGACCAATGATGATGAGTCT 548
DB 112555 GAATGTTTGGCCCTTACCGATGCGACGACTGAGACCAATGATGATGAGTCT 112496

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Osada, N., Kuwada, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and
Hashimoto, K.
isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method
unpublished
2 (bases 1 to 2002)
Hashimoto, K., Osada, N., Kuwada, J. and Sugano, S.
Direct Submission
Submitted (12-APR-2000) Kateuyuki Hashimoto, National Institute of

COMMENT

Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (series)
Osada, N., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and
Hashimoto, K.
Isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method
Unpublished
2 (bases 1 to 2002)
Hashimoto, K., Osada, N., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashimoto@nih.go.jp, URL: <http://www.nih.go.jp/yoken/genbank/>,
Tel:81-3-5285-1111(ext. 2120), Fax:81-3-5285-1181)
URL: <http://www.nih.go.jp/yoken/genbank/>
Lib Name: Sugano mouse brain mncb
Toplo
Vector: pME185-FL3
1st strand cDNA was primed with an oligo(dT) primer
(ATGCGCCCTTTTCTTTTCTTTT); double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5Kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites can be used to isolate the cDNA insert. Library
The DraIII sites can be used to isolate the cDNA insert. Library
was constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing (5' end primer
(CTTCTGCTCTTAAGCGTCCG); 3' end primer
(CGACCTCGACTCGACACAC)).
A part of this sequence is reported in AU067167.

FEATURES
SOURCE

```

/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCB-3035"
/sex="female"
/clone_1b="Sugano mouse brain mncb"
/dev_stage="adult"
637..1827
CDS

```

BASE COUNT	518 a	589 c	516 g	379 f
ORIGIN				

Query Match	31.9%	Score 194;	DB 10;	Length 2002;
Best Local Similarity	99.6%	Pred. No. 3e-103;		
Matches 244;	Conservative	0;	Mismatches 1;	Indels 0;
			Gaps	0;

22 AACCCCTACACCAATGGCTGGAATTAAATCCAGTTGTGGCGCGTCTACAGCCCGAC 81

Db 1240 AACCCCTACACCAATGGCTGGAATTAAATCCAGTTGTGGGGCGGTCTACAGCCCCGAC 1299

82 TTTCTATGCAGGCACCGTGTCTGTGTGTCCAGGCCAACGAGGAGTCTTCCATGTACAGT 141

DB 1300 TTTAAAGCAGGACGGTGTGTTGTTGTCAGGCTAACTCAGGAGGATCTTCTAATGATCAGT 1339

201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 117

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Db 1420 GCTGACACTGCATACGAGGGGCTACCTTCGAGGCCGTTGGTCCGACCGTGTACAAACACC 1479

Qy 262 TTCAG 266
|||||

Db 1480 TTCAG 1484

RESULT 4	AF191501	LOCUS	AF191501	1586 bp	mRNA	linear	ROD 01-MAR-2002
DEFINITION	Mus musculus hexaribonucleotide binding protein 1 (Hrbp1) mRNA, complete cds.						

SOURCE ORGANISM

REFERENCE AUTHORS

JOURNAL

```

FEATURES
source      location/Qualifiers
1. .1586    /location="Misc miscs"just

```

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
1..1586
/gene="Hrbp1"
166..1419
CDS

```

BASE COUNT	407 a	458 c	374 g	347 t
ORIGIN				

Query Match	26.4%	Score 161;	DB 10;	Length 1586;
Best Local Similarity	100.0%	Pred. No. 1.2e-83;		
Matches 161;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

22 AACCCCTACACCAATGGCTGGAATTAATCCAGTTGTGGCGCGGTCTACAGCCCGAC 81

Db 829 AACCCCTACACCAATGGCTGGAATTAAATCCAGTTGTGGGGCGGCTTACAGCCCCCGAC 888

82 TTCTATGCAGGCACCGTGTGTGTGCCAGGCCAACCGAGGATCTTCCATGTACAGT 14

DB 889 TTTAAGCAGCACGGTGTGTTGTCCTCAAGCCATACCAAGAGGGATCTTCCAATGTAACAAT

Year	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

RESULT 5

LOCUS	AC120661	115860 bp	DNA	linear	HTG 23-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-3AC2 *** SEQUENCING IN PROGRESS ***				

21 unordered pieces.

ACCESSION ACT120661

JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
REFERENCE Scios Inc. (US)
AUTHORS Location/Qualifiers
TITLE 1..1340
JOURNAL /organism="Rattus norvegicus"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN

Query Match 100.0%; Score 609; DB 6; Length 1340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTAATAAAGGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTTG 60
DB 535 ATGACTAATAAAGGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTTG 594
QY 61 GCGCGGCTCTACAGCCCGCACTTTATGAGAGCAGCGTCTGTTGTCAGAGCCAAACAG 120
DB 595 GCGCGGCTCTACAGCCCGCACTTTATGAGAGCAGCGTCTGTTGTCAGAGCCAAACAG 654
QY 121 GAGGATCTTCATGATGAGTGGCCCGCACTTGTATATATCTTCAGAGTGGC 180
DB 655 GAGGATCTTCATGATGAGTGGCCCGCACTTGTATATATCTTCAGAGTGGC 714
QY 181 TTTCATATCCGCGCGCCACCTGCTGAGCTGATACCGAGGGGCTCACTTCAGAGCCGT 240
DB 715 TTTCATATCCGCGCGCCACCTGCTGAGCTGATACCGAGGGGCTCACTTCAGAGCCGT 774
QY 241 GGTCCGACCGTGTACACACCTTACAGAGTGGCGGCGCCCGCAATCCCGGCTAT 300
DB 775 GGTCCGACCGTGTACACACCTTACAGAGTGGCGGCGCCCGCAATCCCGGCTAT 834
QY 301 GCGGAGTGTATGATGAAGCAGCGTATGGAATTAATGCTAGAGGTGTTACGCT 360
DB 835 GCGGAGTGTATGATGAAGCAGCGTATGGAATTAATGCTAGAGGTGTTACGCT 894
QY 361 GATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTACAGTACAGTTACGA 420
DB 895 GATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTACAGTACAGTTACGA 954
QY 421 CGAGTTATGCTGCGAGCCCTTACACACACACTTGTCCAGCCCCCACTTACGCGCTT 480
DB 955 CGAGTTATGCTGCGAGCCCTTACACACACTTGTCCAGCCCCCACTTACGCGCTT 1014
QY 481 GGTCCATGAATGCTTTTGGCCCTTACCGATGCAAGTACAGGATGCTATGAT 540
DB 1015 GGTCCATGAATGCTTTTGGCCCTTACCGATGCAAGTACAGGATGCTATGAT 1074
QY 541 GTGGGCTGCTTCTTCTTCATGAGGCTAGATATACCAAGGGGATACAAACGTTT 600
DB 1075 GTGGGCTGCTTCTTCTTCATGAGGCTAGATATACCAAGGGGATACAAACGTTT 1134
QY 601 GCTCCATAT 609
DB 1135 GCTCCATAT 1143

RESULT 2
AF107204 1363 bp mRNA linear ROD 23-JUN-2000
LOCUS AF107204
DEFINITION Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE A novel protein, A2BP, with RNA binding motif, binds to C-terminal ataxin-2

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1363)
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA

FEATURES
source 1..1363
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
1..1363
gene /gene="A2bp"
67..1203
CDS /gene="A2bp"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AA078292.1"
/db_xref="GI:8671588"
/translation="MAQPVASQAPAPONGIPAEYTAAPHAPAEYTGQTPDHTLN
LYPTQTHSEASADTSATVSGTATOTDDAAPTGOPTOPESTESKOPREKLHVN
IPPRFRDPDLRMFGQFKLIDVELIFNERSKGFYFENSADAPREKLHGT
VVEGRKLEVNNAIRVNTNKTVPTNMGKLVNVAIVGAVSPDYAGTVLLCOMREG
SSWYSGSSLVYTSAMPGFYPAATAAARGAALRGRTVNTFRAPAPPPIPAY
VGAMNAPAPLTDATKRSADVDGLVSLQASIVRGYVNRAPY"
358..375
misc_feature /gene="A2bp"
/note="Region: RNA binding motif RNP-2"
misc_feature /gene="A2bp"
/note="Region: RNA binding motif RNP-1"
BASE COUNT 362 a 396 c 324 g 281 t
ORIGIN

Query Match 31.9%; Score 194; DB 10; Length 1363;
Best Local Similarity 99.6%; Pred. No. 3e-103;
Matches 244; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 22 AACCCCTACACCAATGCTGGAATTAATCCAGTTGGCGCGCTACAGCCCCCAG 81
DB 616 AACCCCTACACCAATGCTGGAATTAATCCAGTTGGCGCGCTACAGCCCCCAG 675
QY 82 TTCTATGAGGACGAGTGTGTTGTCAGGACCAACAGAGAGATCTTCATGTAAGT 141
DB 676 TTCTATGAGGACGAGTGTGTTGTCAGGACCAACAGAGAGATCTTCATGTAAGT 735
QY 142 GCGCCAGTTCACTGTATATATCTTGCATGCTGCTTCATATCCGCGCCACT 201
DB 736 GCGCCAGTTCACTGTATATATCTTGCATGCTGCTTCATATCCGCGCCACT 795
QY 202 GCTGACGCTGCTACCGAGGGGCTCACTTGAAGCCGTGTGCGACCGTGTACAAACC 261
DB 796 GCTGACGCTGCTACCGAGGGGCTCACTTGAAGCCGTGTGCGACCGTGTACAAACC 855
QY 262 TTCAG 266
DB 856 TTCAG 860

RESULT 3
AB041596 2002 bp mRNA linear ROD 30-JUN-2000
LOCUS AB041596
DEFINITION Mus musculus brain cDNA, clone MNCD-3035, similar to Homo sapiens
CDNA FLJ10165, clone HEMBA1003591.
ACCESSION AB041596
VERSION AB041596.1 GI:7670455
KEYWORDS f1s (full insert sequence).
SOURCE Mus musculus (strain:C57BL) adult female cDNA to mRNA,
clone:lib:Sugano mouse brain mncd clone:MNCD-3035.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 14:22:28 ; Search time 1627.96 Seconds

(without alignments)
10887.019 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactataataaaagccgt.....acaacgcttgcctcatat 609Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: GenEmbl.*

```
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ats:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_ats:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rnd:*
36: em_hcg_mam:*
37: em_hcg_vtc:*
38: em_ay:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	609	100.0	1340	6 AX268800	AX268800 Sequence
2	194	31.9	1363	10 AF107204	AF107204 Mus muscu
3	194	31.9	2002	10 AB041596	AB041596 Mus muscu
4	161	26.4	1586	10 AF191501	AF191501 Mus muscu
5	92	15.1	115860	2 AC120661	AC120661 Rattus no
6	79	13.0	73929	2 AC103207	AC103207 Rattus no
7	70	11.5	1475	9 AB060859	AB060859 Macaca fa
8	70	11.5	1513	9 AK001027	AK001027 Homo sapi
9	70	11.5	2000	9 AF229057	AF229057 Homo sapi
10	70	11.5	2279	9 AF107203	AF107203 Homo sapi
11	70	11.5	2372	6 AR134676	AR134676 Sequence
12	70	11.5	2971	9 HSFOX14	AF109120 Homo sapi
13	70	11.5	3348	9 AF109106	AF109106 Homo sapi
14	70	11.5	164538	9 AC005774	AC005774 Homo sapi
15	67	11.0	115860	2 AC120661	AC120661 Rattus no
16	54	8.9	1547	9 AF094849	AF094849 Homo sapi
17	43	7.1	374	9 HSFOX08	AF109114 Homo sapi
18	43	7.1	1623	9 HSM803066	AF171700 Homo sapi
19	40	6.6	234	9 HSFOX13	AF109119 Homo sapi
20	26	4.3	251	9 HSFOX09	AF109115 Homo sapi
21	26	4.3	596	9 HSA323668	AJ323668 Homo sapi
22	26	4.3	617	9 HSA335811	AJ335811 Homo sapi
23	26	4.3	653	9 HSA335616	AJ335616 Homo sapi
24	26	4.3	675	9 HSA334013	AJ334013 Homo sapi
25	23	3.8	151	9 HSFOX11	AF109117 Homo sapi
26	22	3.6	154471	9 AC022523	AC022523 Homo sapi
27	22	3.6	184497	2 AC099688	AC099688 Homo sapi
28	22	3.6	210713	2 AC024271	AC024271 Homo sapi
29	21	3.4	105589	2 AC092269	AC092269 Homo sapi
30	21	3.4	157613	9 AC018769	AC018769 Homo sapi
31	21	3.4	166847	9 AC008393	AC008393 Homo sapi
32	21	3.4	180520	2 AC113426	AC113426 Homo sapi
33	21	3.4	186872	9 AC074052	AC074052 Homo sapi
34	21	3.4	188046	2 AC124857	AC124857 Homo sapi
35	21	3.4	195068	9 AC012100	AC012100 Homo sapi
36	21	3.4	197000	2 AC124859	AC124859 Homo sapi
37	21	3.4	207751	2 AC112189	AC112189 Homo sapi
38	20	3.3	249	9 HSFOX10	AF109116 Homo sapi
39	20	3.3	4684	9 AK001678	AK001678 Homo sapi
40	20	3.3	6123	2 AF303002	AF303002 Homo sapi
41	20	3.3	91531	2 AC079846_3	Continuation (4 of
42	20	3.3	116539	2 AC123239	AC123239 Rattus no
43	20	3.3	118369	2 AC131018	AC131018 Rattus no
44	20	3.3	139242	10 AL590996	AL590996 Mouse DNA
45	20	3.3	140186	2 AC055808	AC055808 Homo sapi

ALIGNMENTS

RESULT 1
AX268800
LOCUS AX268800 1340 bp
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Stanton, L.W. and White, R.T.
TITLE Secreted factors

THIS PAGE BLANK (USPTO)

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 454
OTHER INFORMATION: n = A,T,C or G
US-09-998-598-2571

Query Match 5.1%; Score 31; DB 10; Length 466;
Best Local Similarity 57.9%; Pred. No. 5.5;
Matches 55; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 166 TCTGCAATGCTGCTTTCATATCCGCCGCTGCTGCACTGCATATCCGAGGGCT 225
DB 397 TTGCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338
QY 226 CACCTTCGAGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 260
DB 337 GACCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 303

Search completed: March 15, 2003, 16:12:27
Job time : 564.057 secs

APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 7757
LENGTH: 927
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(927)
US-09-815-242-7757

Query Match 5.1%; Score 31.2; DB 10; Length 927;
Best Local Similarity 49.4%; Pred. No. 6.6;
Matches 81; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 341 TGCTACAGGAGTGTACGCTGCATACCGCTACGCCGCCACCCCTGCACCTGCTG 400
DB 395 TGCTCAGCAGAGTGTCTGTGGCGCGCAGCAGCGCGGACAGTGTGCGGCGG 336
QY 401 CCTACAGTACGACTTACGACGAGTTATGCTGCCACCCCTACGACACACTTGCTC 460
DB 335 GGCACCATATGACGATGCGCTTGTGGCTGTCTGTGACACGATGACGCGCACTCGCTG 276
QY 461 CAGCCCCCAGCTACGCGGCTGTGTCATGATGCTTTTGCGGCC 504
DB 275 GACTCCCCGCGCACCCGCTGACGCCCTGCGACTGCATGCCCTC 232

RESULT 13
US-09-557-232-3/C
Sequence 3, Application US/09557232
Publication No. US20030040099A1
GENERAL INFORMATION:
APPLICANT: Frey, Terry
APPLICANT: Pougatchev, Konstantin
APPLICANT: Abernathy, Emily
TITLE OF INVENTION: Highly Infectious Rubella Virus Clones and Methods of Productio
FILE REFERENCE: 07362-0104/43439-221816
CURRENT APPLICATION NUMBER: US/09/557,232
CURRENT FILING DATE: 2000-04-24
PRIOR APPLICATION NUMBER: 08/459,041
PRIOR FILING DATE: 1995-06-02
PRIOR APPLICATION NUMBER: 08/093,453
PRIOR FILING DATE: 1993-07-19
PRIOR APPLICATION NUMBER: 07/722,334
PRIOR FILING DATE: 1991-06-28
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent version 3.0
SEQ ID NO: 3
LENGTH: 2558
TYPE: DNA
ORGANISM: Rubella virus

US-09-557-232-3

Query Match 5.1%; Score 31.2; DB 9; Length 2558;
Best Local Similarity 60.7%; Pred. No. 11;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 168 TGCAATGCTGCTTTCATATCCGGCCGACATGCTGCACCTGCATACCGAGGGCTCA 227
DB 587 TGCCATGTCACCGCGGAGGACCATATGTCATCCGACATGCTCGCGGGGCTGA 528
QY 228 CCTCGAGGCGGTGATGACCGT 251
DB 527 CTTCCGCGGCTTGAGCACACTGT 504

RESULT 14
US-09-557-232-1/C
Sequence 1, Application US/09557232
Publication No. US20030040099A1
GENERAL INFORMATION:
APPLICANT: Frey, Terry
APPLICANT: Pougatchev, Konstantin
APPLICANT: Abernathy, Emily
TITLE OF INVENTION: Highly Infectious Rubella Virus Clones and Methods of Productio
FILE REFERENCE: 07362-0104/43439-221816
CURRENT APPLICATION NUMBER: US/09/557,232
CURRENT FILING DATE: 2000-04-24
PRIOR APPLICATION NUMBER: 08/459,041
PRIOR FILING DATE: 1995-06-02
PRIOR APPLICATION NUMBER: 08/093,453
PRIOR FILING DATE: 1993-07-19
PRIOR APPLICATION NUMBER: 07/722,334
PRIOR FILING DATE: 1991-06-28
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent version 3.0
SEQ ID NO: 1
LENGTH: 9759
TYPE: DNA
ORGANISM: synthetic construct - rubella
US-09-557-232-1

Query Match 5.1%; Score 31.2; DB 9; Length 9759;
Best Local Similarity 60.7%; Pred. No. 20;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 168 TGCAATGCTGCTTTCATATCCGGCCGACATGCTGCACCTGCATACCGAGGGCTCA 227
DB 3389 TGCCATGCCAGCGCGGAGGACCATATGTCATCCGACATGCTCGCGGGGCTGA 3330
QY 228 CCTCGAGGCGGTGATGACCGT 251
DB 3329 CTTCCGCGGCTTGAGCACACTGT 3306

RESULT 15
US-09-998-598-2571/C
Sequence 2571, Application US/09998598
Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO: 2571
LENGTH: 466
TYPE: DNA


```

RESULT 7
US-10-051-643-180
: Sequence 180, Application US/10051643
: Publication No. US20020197265A1
: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Tan, Paul L. J.
: TITLE OF INVENTION: Methods and Compounds for the Treatment
: TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
: TITLE OF INVENTION: System using Mycobacterium Vaccae
: FILE REFERENCE: 11000.1008C2
: CURRENT APPLICATION NUMBER: US/10/051,643
: CURRENT FILING DATE: 2002-01-18
: PRIOR APPLICATION NUMBER: US09/156,181
: PRIOR FILING DATE: 1998-09-17
: PRIOR APPLICATION NUMBER: US 08/996,624
: PRIOR FILING DATE: 1997-12-23
: NUMBER OF SEQ ID NOS: 208
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 180
: LENGTH: 1071
: TYPE: DNA
: ORGANISM: Mycobacterium vaccae
US-10-051-643-180

```

Query Match	Similarity	5.3*	Score 32.2;	DB 9;	Length 1071;
Best Local	Similarity	52.6*	Pred. No. 3.4;		
Matches	70;	Conservative	0;	Mismatches	63;
				Indels	0;
				Gaps	0;
QY	346	CAGGGTGGTTACGCTGTACATCCGCTACGCCACCCCTGTGCACATGCTGCTGTAC	405		
Db	102	CATCGCGGTCAACAGGGGGTGGCCCGCGGCGCTCCCGCAGCCCCGCGCCCC	161		
QY	406	AGTGAAGTTACGACGAGATTATATCTGCGCACCCCTTACACACACACTTGCTCCAGCC	465		
Db	162	TGCCACAGTGAAGCAAAACCGTCACGGTTGGCGCCCAAGCGCGGCCACAACTCATCCGCG	221		
QY	466	CCCACCTACGGCG	478		
Db	222	CCCCGGTGTGACG	234		

RESULT 9
US-09-738-626-3418
; Sequence 3418, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO

Db 317 SSCASSCWS.SCSHCYSCSSCCSSCCSSCCSSCCSSCCSSCCSNHNCSC 258
Qy 323 CAGTGTATGCAATAATGTTACAGGGTGTACCTGATACCGTACGCCAGCCCA 382
Db 257 SCSSES.SCSCCSSSSCCSSCCSS...NAGC.N.NSTYBSCR..TBTC.NMC.N.CCN 198
Qy 383 CCCCTGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 442
Db 197 CC.CW.TC..NKC..SCTBNC.RCDBM.AC.CC.NC.C.CC.CC..SC..CPCSA.WTCS 138
Qy 443 ACCACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
Db 137 C.BC.WSCYBTHTCY.BH.T.YC..DC 110

RESULT 4

US-10-184-644-88/c
Sequence 88, Application US/10184644
Publication No. US20030044930A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C227
CURRENT APPLICATION NUMBER: US/10/184,644
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 88
LENGTH: 735
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-644-88

Query Match 5.6%; Score 34; DB 9; Length 735;
Best Local Similarity 7.6%; Pred. No. 0.77; Mismatches 172; Indels 0; Gaps 0;
Matches 23; Conservative 109; Mismatches 172; Indels 0; Gaps 0;

Qy 61 GCGCGGTCTACAGCCCGCACTTTCTATGACGACGAGTGTGTCAGGCCAAGCAG 120
Db 307 RBCSB..TNHDSM..MKMYH..D.SA..H.SBSGMKHNMB.BGB.B.Y.N..YR. 248
Qy 121 GAGGATTTCCATGATACAGTGGCCCACTTCTATGATCTTCCATGCTGCTG 180
Db 247 MHBNT...Y.MB..HMC.Y..YHNTB..B..BRMATM.A.YMDYTMW.S...BNM 188
Qy 181 TTTCATATCGCGCGGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 187 TT.N.ANDSGCSCTBSM.NMT..MRYNATSMK..BRS.N..B..CSGAS.SBTS 128
Qy 241 GGTGCACTGATACACACCTTCAAGCTGCGGCGCCCGCCCGCCCGCCCGCTAT 300
Db 127 HSRCTBDBDRGRDCAANT.SBHACH..RDA.A.SST..C.NY..N..MS.Y..Y.N. 68
Qy 301 GCGGAGTAGTATACAGACGCACTGATGCGCAATTAATGCTACAGGCTGCTTAC 360
Db 67 B..DNMSH.SMB..W.HCSYSTSB..HTYC.NW.SBCTT.G.T..TCT.T..TYT.SB 8
Qy 361 GCAT 364
Db 7 ...Y 4

RESULT 5
US-10-184-644-402/c
Sequence 402, Application US/10184644
Publication No. US20030044930A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C227
CURRENT APPLICATION NUMBER: US/10/184,644
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 402
LENGTH: 653
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-644-402

Query Match 5.3%; Score 32.2; DB 9; Length 653;
Best Local Similarity 5.4%; Pred. No. 2.7; Mismatches 311; Indels 0; Gaps 0;
Matches 27; Conservative 158; Mismatches 311; Indels 0; Gaps 0;

Qy 100 CTGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 159
Db 588 H...SSSS.GYHM.M..AM.R..YH.MC..M.B..MDN..STISANT.B.RG.HMTN.H 529
Qy 160 TATCTTCTGCAATGCTGCTGCTTTCATATCCGCGCGCACTGCTGCACTGCA 219
Db 528 ..NSTYSDYSSYS.SYVYMDMDYDCSSSSSSSSSSSS..HTSSBSSSSAAMYMM 469
Qy 220 GGGCTACCTTGAGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 279
Db 468 YM..M..MY.M.TSA..MMAM.M..M.T.M.Y..SM.KMD.M..TH.THM.A.H.T. 409
Qy 280 CCACCCCAATCCGCGGCTATGCGGAGTAGTATGACAGCAGCAGTATGCAATAA 339
Db 408 TMRRS..TN..M.....CCY.B.A.G.RMYTNYDA.GN..T..H..TMN.SCMATR.T 349
Qy 340 TTGCTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 399
Db 348 CYTB.T.BN.MH...TMNR.NKTHDYCBM.RHBC.MBGM.TWSTS..MMY.TST.HH.S. 289
Qy 400 GCTTACGTCAGTATGAGCAGGCTTATGCTGCCAGCCCTTACACACACACTGCT 459
Db 288 NMS..CYK.S..NS.H.C.M.....BBC.NB.C.SS.KBN.RS..SNSN..BSYRR... 229
Qy 460 CCAGCCCACTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 519
Db 228 ..SS.BC.M.CS.D..SSSR.SBT.M.DRYN.HM.CTY..HCA.RRS..HCDSDS.BH 169
Qy 520 ACTAGAGCATGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 579
Db 168 Y..G.TA...DT.HYK.CSC.G..BK..C..Y.SS.YC.B.HCY..HY...YHKS... 109
Qy 580 CAAAGGAGTACAAC 595
Db 108 K.....K.TRDH.SH 93

RESULT 6
US-09-738-626-3417/c

301 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATAATGCTACAGGGGTGTTACGCT 360
835 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATAATGCTACAGGGGTGTTACGCT 894
361 GCATACCGCTACGCGCCAGCCACCCCTGCGCACTGCTGCTGCTACAGTACAGTTACGGA 420
895 GCATACCGCTACGCGCCAGCCACCCCTGCGCACTGCTGCTGCTACAGTACAGTTACGGA 954
421 CGAGTTATGCTGCGCGACCCCTACACACACACTGCTGCTGCGCGCGCGCGCGCGCTT 480
955 CGAGTTATGCTGCGCGACCCCTACACACACACTGCTGCTGCGCGCGCGCGCGCGCTT 1014
481 GGTGCGCATGATGCTTTTGGCGCTGACCGATGCGCAAGTACAGAGCCATG-ATGATGAT 540
1015 GGTGCGCATGATGCTTTTGGCGCTGACCGATGCGCAAGTACAGAGCCATG-ATGATGAT 1074
541 GTGGCTCTGCTTTCTTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGCTTTT 600
1075 GTGGCTCTGCTTTCTTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGCTTTT 1134
601 GCTCCATAT 609
1135 GCTCCATAT 1143

RESULT 2

US-09-794-591-1
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulse, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patencin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987) .. (1979)
US-09-794-591-1

Query Match 70.4%; Score 428.6; DB 10; Length 2372;
Best Local Similarity 82.0%; Pred. No. 2.1e-125;
Matches 542; Conservative 0; Mismatches 64; Indels 55; Gaps 2;

1 ATGACTAATAAAGGCGCGTGAACCCCTACACCAATGGCTGAATAATTCAGTTGTG 60
1572 ATGACAAATAAAGGCGCGTGAACCCCTACACCAATGGCTGAATAATTCAGTTGTG 1631
61 GCGCGGCTACAGCCCGCATTTCTATGAGGACCGGTGCTGTGTGCCAGGCAACAG 120
1632 GCGCGGCTACAGCCCGCATTTCTATGAGGACCGGTGCTGTGTGCCAGGCAACAG 1691
121 GAGGATCTTCATGTAAGTGGCGCCAGTTCACTTGTATATCTTGTGAATGCTGCGC 180
1692 GAGGATCTTCATGTAAGTGGCGCCAGTTCACTTGTATATCTTGTGAATGCTGCGC 1751
181 TTTCATATCCGCGCGCACTGCTGAGTGCATACGAGGGGCTCACTTCAGAGCGCT 240
1752 TTTCATATCCGCGCGCACTGCTGAGTGCATACGAGGGGCTCACTTCAGAGCGCT 1811
241 GGTGCAACCGTGTACACACTTTCAGAGTGGCGCGCGCGCGCGCGCGCGCGCTAT 300

1812 GGTGCAACCGTGTACACACTTTCAGAGTGGCGCGCGCGCGCGCGCGCGCGCTAC 1871
301 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATAATGCTACAGGGGTGTTACGCT 360
1872 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATAATGCTACAGGGGTGTTACGCT 1929
361 GCATACCGCTACGCGCCAGCCACCCCTGCGCACTGCTGCTGCTACAGTACAGTTACGGA 420
1930 GCATACCGCTACGCGCCAGCCACCCCTGCGCACTGCTGCTGCTACAGTACAGTTACGGA 1989
411 -----CAGTTACGAGAGTTT 427
1990 TTGCTCTGCTTGTGAGAGATGAAATTTCTTGAACACTCTGCTGCTGCTGCTGCTT 2049
428 ATGTCGCGAGCCCGCTACACCAACACACTTGTGCTGCGCGCGCGCGCGCGCGCTGCGCA 487
2050 ATGTCGCGAGCCCGCTACACCAACACACTTGTGCTGCGCGCGCGCGCGCGCGCTGCGCA 2109
488 TGAATGCTTTTGGCGCGCTTGAACCGATGCGCAAGTACAGAGCCATGCTATGATGTCGTC 547
2110 TGAATGCTTTTGGCGCGCTTGAACCGATGCGCAAGTACAGAGCCATGCTATGATGTCGTC 2169
548 TCGTCTCTTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGCTTTGCTCCAT 607
2170 TCGTCTCTTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGCTTTGCTCCAT 2229
608 A 608
2230 A 2230

RESULT 3

US-10-184-644-202/C
; Sequence 202, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Aubin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 202
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-202

Query Match 5.8%; Score 35.4; DB 9; Length 440;
Best Local Similarity 15.9%; Pred. No. 0.22;
Matches 52; Conservative 88; Mismatches 188; Indels 0; Gaps 0;

143 GCGCCAGTTCACTTGTATATCTTCTGCAATGCTGCGCTTTCCATATCGGCGCGCACATG 202
437 SSV.HMN.THMN..C.MS.NMW.AH.N.KC.SA.SNBANBCCBTHCCSCWSSC.CYN 378
203 CTGACAGTGCATACGAGGGGCTCACTTTCAGAGCGCGGTGTCGACACGCTGTAACACCT 262
377 HCSGC..YMC.MS..YKSSBC.CY.C..CS.CSCTT.NC.M.GC.MDNCSCSCDNC 318
263 TCAGAGCTGCTAT 322

GenCore version 5.1.4.g5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:44:58 ; Search time 54.057 Seconds
(without alignments)
7907.139 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactataataaaagccgct.....acaacgcttcctccatcat 609

Scoring table: Gapped 10'-0, Gapext 1.0

Searched: 501302 seqs, 350932545 residues

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTIS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	609	100.0	1340	US-09-809-545A-1	Sequence 1, Appl1
2	428.6	70.4	2372	US-09-794-591-1	Sequence 1, Appl1
3	35.4	5.8	440	US-10-184-644-202	Sequence 202, App
4	34	5.6	735	US-10-184-644-88	Sequence 88, Appl
5	32.2	5.3	653	US-10-184-644-402	Sequence 402, App
6	32.2	5.3	951	US-09-738-626-3417	Sequence 3417, App
7	32.2	5.3	1071	US-10-051-643-180	Sequence 180, App
8	32.2	5.3	1071	US-09-880-505-180	Sequence 180, App
9	32.2	5.3	1185	US-09-738-626-3418	Sequence 3418, App
10	32.2	5.3	3309400	US-09-738-626-1	Sequence 1, Appl1
11	31.2	5.1	1653	US-09-897-214-9	Sequence 9, Appl1
12	31.2	5.1	927	US-09-815-442-7757	Sequence 7757, App
13	31.2	5.1	2558	US-09-557-232-3	Sequence 3, Appl1
14	31.2	5.1	9759	US-09-557-232-1	Sequence 1, Appl1
15	31	5.1	466	US-09-998-598-2571	Sequence 2571, App
16	31	5.1	1572	US-09-908-928-3	Sequence 3, Appl1
17	31	5.1	2127	US-09-908-928-1	Sequence 1, Appl1
18	31	5.1	15720	US-10-025-380-1058	Sequence 1058, App
19	31	5.1	15720	US-09-922-217-1058	Sequence 1058, App

C	20	31	5.1	15720	10	US-09-833-263-1058	Sequence 1058, App
C	21	30.6	5.0	671	9	US-10-184-644-346	Sequence 346, App
C	22	30.6	5.0	1459	9	US-09-964-824A-277	Sequence 277, App
C	23	30.4	5.0	678	9	US-10-184-644-150	Sequence 150, App
C	24	30.4	5.0	1141	9	US-10-184-644-120	Sequence 120, App
C	25	30.4	5.0	3642	10	US-09-969-515-7	Sequence 7, Appl1
C	26	30.4	5.0	3675	10	US-09-969-515-1	Sequence 1, Appl1
C	27	30.4	5.0	4666	10	US-09-969-515-13	Sequence 13, Appl1
C	28	30.2	4.9	750	9	US-10-184-644-104	Sequence 104, App
C	29	30.2	4.9	403	10	US-09-960-352-8222	Sequence 8222, App
C	30	30	4.9	1873	9	US-09-986-480-94	Sequence 94, Appl1
C	31	30	4.9	3639	9	US-10-012-896-779	Sequence 779, App
C	32	30	4.9	3639	9	US-09-895-793-779	Sequence 779, App
C	33	30	4.9	3639	9	US-09-895-814-779	Sequence 779, App
C	34	30	4.9	3639	10	US-09-759-143-779	Sequence 779, App
C	35	30	4.9	3639	10	US-09-780-669-779	Sequence 779, App
C	36	30	4.9	3639	10	US-09-822-827-779	Sequence 779, App
C	37	30	4.9	3918	10	US-09-781-311-1	Sequence 1, Appl1
C	38	30	4.9	5668	9	US-10-012-896-777	Sequence 777, App
C	39	30	4.9	5668	9	US-09-895-793-777	Sequence 777, App
C	40	30	4.9	5668	9	US-09-895-814-777	Sequence 777, App
C	41	30	4.9	5668	10	US-09-759-143-777	Sequence 777, App
C	42	30	4.9	5668	10	US-09-780-669-777	Sequence 777, App
C	43	30	4.9	5668	10	US-09-822-827-777	Sequence 777, App
C	44	30	4.9	11869	10	US-09-764-869-2282	Sequence 2282, App
C	45	29.8	4.9	360	10	US-09-867-550-519	Sequence 519, App

ALIGNMENTS

RESULT 1									
US-09-809-545A-1									
Sequence 1, Application US/09809545A									
Patent No. US20020110804A1									
GENERAL INFORMATION: Stanton, Lawrence W.									
APPLICANT: Stanton, Lawrence W.									
TITLE OF INVENTION: SECRETED FACTORS									
FILE REFERENCE: SCIOS.017A									
CURRENT APPLICATION NUMBER: US/09/809,545A									
CURRENT FILING DATE: 2001-03-14									
NUMBER OF SEQ ID NOS: 84									
SOFTWARE: FastSeq for Windows Version 4.0									
SEQ ID NO 1									
LENGTH: 1340									
TYPE: DNA									
ORGANISM: Rattus norvegicus									
US-09-809-545A-1									
Query Match									
Best Local Similarity 100.0%; Pred. No. 1.4e-182;									
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	ATGACTAATAAAGCCGTGACACCCCTACACCAATGCTGAAATTAAATCCAGTTTG	60						
DB	535	ATGACTAATAAAGCCGTGACACCCCTACACCAATGCTGAAATTAAATCCAGTTTG	594						
QY	61	GGCGGGCTACAGCCCGGACTTTCATGACGACGCGGTGTGTCAGAGCAACAG	120						
DB	595	GGCGGGCTACAGCCCGGACTTTCATGACGACGCGGTGTGTCAGAGCAACAG	654						
QY	121	GAGGATTCATGATGACAGTGGCCCGGAGTTCATGATGATGATGATGATGATG	180						
DB	655	GAGGATTCATGATGACAGTGGCCCGGAGTTCATGATGATGATGATGATGATG	714						
QY	181	TTTCCATATCCGCGCCGCTGCTGACGCTGATGATGATGATGATGATGATG	240						
DB	715	TTTCCATATCCGCGCCGCTGCTGACGCTGATGATGATGATGATGATGATG	774						
QY	241	GATGACCGGTGATGACACCTTTCAGAGCTGCGGCGCCCAACCCCAATCCGCGCAT	300						
DB	775	GATGACCGGTGATGACACCTTTCAGAGCTGCGGCGCCCAACCCCAATCCGCGCAT	834						

THIS PAGE BLANK (USPTO)

RESULT 14

US-08-997-080-180

; Sequence 180, Application US/08997080

; Patent No. 5968524

; GENERAL INFORMATION:

; APPLICANT: WATSON, JAMES D.

; APPLICANT: TAN, PAUL E.J.

; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-

; NUMBER OF SEQUENCES: 194

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Law Offices of Ann W. Speckman

; STREET: 2601 Elliott Avenue, Suite 4185

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98121

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/997,080

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sleath, Janet

; REGISTRATION NUMBER: 37,007

; REFERENCE/DOCKET NUMBER: 11000.1007

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-269-0565

; TELEFAX: 206-269-0563

; TELEX:

; INFORMATION FOR SEQ ID NO: 180:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1071 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-997-080-180

Query Match 5.3%; Score 32.2; DB 2; Length 1071;
Best Local Similarity 52.6%; Pred. No. 2.2;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 346 CAGGATGTTACGCTGATACCGCTACGCCAGCCACCCCTGCACTGCTGCTTAC 405
DB 102 CATCGCGTCCACACGCGGTGCGCCAGCCGCGCTCTCCGACGCCCTGCGCGCC 161

QY 406 AGTACAGTTACGACGAGATTATGCTGCCGACCCCTTACCAACACACTTGTCCAGCC 465
DB 162 TGCCACAGTACGCAACCGTACGCTTGCGCCCAACGCGCGGCAACAATCATCCCGCG 221

QY 466 CCCACCTACGCGC 478
DB 222 CCGCGTGTGACG 234

RESULT 15

US-08-997-362-180

; Sequence 180, Application US/08997362

; Patent No. 5985287

; GENERAL INFORMATION:

; APPLICANT: Tan, Paul

; APPLICANT: Hiyama, Jun

; APPLICANT: Visser, Elizabeth

; APPLICANT: Skinner, Margot

; APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS

; NUMBER OF SEQUENCES: 194

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Law Offices of Ann W. Speckman

; STREET: 2601 Elliott Avenue, Suite 4185

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98121

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/997,362

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970

; FILING DATE: June 12, 1997

; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347

; FILING DATE: August 29, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Sleath, Janet

; REGISTRATION NUMBER: 37,007

; REFERENCE/DOCKET NUMBER: 11000.1002c2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-269-0565

; TELEFAX: 206-269-0563

; TELEX:

; INFORMATION FOR SEQ ID NO: 180:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1071 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-997-362-180

Query Match 5.3%; Score 32.2; DB 2; Length 1071;
Best Local Similarity 52.6%; Pred. No. 2.2;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 346 CAGGATGTTACGCTGATACCGCTACGCCAGCCACCCCTGCACTGCTGCTTAC 405
DB 102 CATCGCGTCCACACGCGGTGCGCCAGCCGCGCTCTCCGACGCCCTGCGCGCC 161

QY 406 AGTACAGTTACGACGAGATTATGCTGCCGACCCCTTACCAACACACTTGTCCAGCC 465
DB 162 TGCCACAGTACGCAACCGTACGCTTGCGCCCAACGCGCGGCAACAATCATCCCGCG 221

QY 466 CCCACCTACGCGC 478
DB 222 CCGCGTGTGACG 234

Search completed: March 15, 2003, 16:22:18
Job time: 1277.68 secs

LOCATION: 15429.15664
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: /frame"
US-08-402-068-3

Query Match 5.6%; Score 34.2; DB 1; Length 15664;
Best Local Similarity 47.8%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 220 GGGGCTCAGCTTCGAGCGGCTGTCGACCGTGTACAAACCTTCAGAGTGGCGCGCC 279
Db 720 GGGGCTCAGCTTCGAGCGGCTGTCGACCGTGTACAAACCTTCAGAGTGGCGCGCC 779
Qy 280 CCACCCCAATCCCGGCTATGGCGGAGTGTATCAAGAGCAAGTATGCAATAA 339
Db 780 GCCGTGCTATGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
Qy 340 TTGCTACAGGCTGTATGCTGCAATACCGCTACGCCCGCGCGCGCGCGCGCGCGCT 399
Db 840 CTGACTTGTGCGGAGCGCGCGGTTGGCTATGAGCGCGCGCGCGCGCGCGCGCGCA 899
Qy 400 GCCTACAGTACAGTACGAGAGATT 426
Db 900 GCCACACCGTGTGCCAAGCCCGCGAT 926

RESULT 10

US-09-056-180/c
Sequence 180, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 180:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-180

TREATM

Query Match 5.5%; Score 33.4; DB 4; Length 538;
Best Local Similarity 51.0%; Pred. No. 0.68;
Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 357 CGGTGATACCGGTACCGCCAGCCCGCTGCGACTGCTGCTCACTGACAGTGA 416
Db 143 CGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 109

Db 263 CGTACCGCGGTACCGCCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 204
Qy 417 CGGACGATTATGCTGCGGACCCCTTACCAACACACTTGTCTACGCGCGCGCGCGCG 476
Db 203 CGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 144
Qy 477 CGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 511
Db 143 CGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 109

RESULT 11

US-09-072-596-175/c
Sequence 175, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 175:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-175

Query Match 5.5%; Score 33.4; DB 4; Length 538;
Best Local Similarity 51.0%; Pred. No. 0.68;
Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 357 CGGTGATACCGGTACCGCCAGCCCGCTGCGACTGCTGCTCACTGACAGTGA 416
Db 263 CGTACCGCGGTACCGCCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 204
Qy 417 CGGACGATTATGCTGCGGACCCCTTACCAACACACTTGTCTACGCGCGCGCGCGCG 476
Db 203 CGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 144
Qy 477 CGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 511
Db 143 CGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 109

CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSER: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,068
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fugitt, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding"
OTHER INFORMATION: sequence
OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1115..11786
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature

```
NAME/KEY: misc_feature
LOCATION: 747..1109 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376 /function= "potential coding
OTHER INFORMATION: sequence"
OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9450..10244 /function= "potential open reading
OTHER INFORMATION: frame"

OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11115..11786 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664 /function= "potential open reading
OTHER INFORMATION: frame"
US-08-402-066-3

Query Match 5.6%; Score 34.2; DB 1; Length 15664;
Best Local Similarity 47.8%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 220 GGGGCTACCTTGAGGGCGTGTGACCGCTGATACACCTTGAGAGCTGGCGGCC 279
DB 720 GCGCGTCTACTGGAGGCGGTGCTGCATGATCCGCGACATGCGAGCTGGCC 779
QY 280 CCACCCCAATCCGCGCTATGCGGAGTAGTATCAAGCCAGTATGGCAATAA 339
DB 780 GCGGTGCTATCGCGCGCGCCGCCCGGCGGCAACCGCAACAGAGCGGCGTTT 839
QY 340 TTGCTACAGGGGTGTTACCTGATACCGCTACGCCCGCACCCCTGCGACTGCT 399
DB 840 CTGTACTTCTCGCGGAGCGCGGTTGCGCTATATGACGCGCGCGCTATGCTCCA 899
QY 400 GCGTACAGTGACAGTTACGAGCGAGTT 426
DB 900 GCGCACACCGTGTGCCAAGCCCGGAT 926

RESULT 9
US-08-402-068-3
Sequence 3, Application US/08402068
Patent No. 5633159
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
NUMBER OF SEQUENCES: 6
```

OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15425..15664
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
US-08-508-004-3

Query Match 5.6%; Score 34.2; DB 1; Length 15664;
Best Local Similarity 47.8%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 220 GGGGCTACCTTCGAGGCGGTGTCGACCGTGTATCAACACCTTCAGAGCTGCGCGCC 279
DB 720 GGGGCTACTGAGAGGCGGTGCTGTCATGATCCGCGAGCAATTCGAGGCTGCGC 779
QY 280 CCACCCCAATCCCGGCTATGCGGAGTGTATCAAGACCAAGTATGCGCAATAA 339
DB 780 GCGGTGCTATGCGGCG 839
QY 340 TTGCTACAGGCTGTATGCTGTCATACCGCTACCGCGCGCGCGCGCGCGCGCGCT 399
DB 840 CTGTACTTGTCTCGGCGAGCGCGGTTCGCTATGAGCAGCGCGCGCGGTATGTCGCA 899
QY 400 GCCTACAGTGAAGTTACGAGCGAGTT 426
DB 900 GGCACACCGTGTGCCAAGCCCGCAT 926

RESULT 8
US-08-402-066-3
Sequence 3, Application US/08402066
Patent No. 5612182
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/402,066
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Puglt, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:

```
NAME/KEY: misc_feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
US-08-402-282-3
```

```
Query Match 5.6%; Score 34.2; DB 1; Length 15664;
Best Local Similarity 47.8%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
Qy 220 GGGGTCACCTTCAGGCGCGTGTGCGACCGGTGTCACACCTTACAGGCTGCGCGCC 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 720 GCGCGTCTACCTGAGGCGGCGGTGCTGTCATGATCCGCGACGACATCGGCGTGGCC 779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 280 CCACCCCAATCCGCGCGTATGCGAGTAGTATCAAGACGAGTATGCGCAATAA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 780 GCGGTGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 340 TTGCTACAGGAGGTGTTACGCTGATACCGCTACGCCGACGCCCGCGCGCGCTGCT 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 840 CTGACTGCTGCGGAGGCGCGGTTCGCTATGACGAGCGCGCGCGCTGATGCTCGCA 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 400 GCCTACAGTACAGTTACGAGCAGTT 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 900 GGCACACCGGTGTGCCAAGCCCGCGAT 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 7
US-08-508-004-3
Sequence 3, Application US/08508004
Patent No. 5582969
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,004
FILING DATE: 27-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,282
FILING DATE: 10-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
sequence"
```


QY 404 ACAGTACAGTACGACAGGTTATGCTGCCGACCCCTACACACACACTTCTCCAG 463
Db 3942479 ATGCAGACACCCCGCGCGCTTGTCCGCCGCGCGCGCGCGGACCGGTG 3942420
QY 464 CCCCCACCTACGCGCTTGTGTCATGATGCTTGGCGCTTGACCATGCGCAAGACTA 523
Db 3942419 CCGCGCGGTGCGCGCTTGTCCGCCGCGCGCGCGCGCGCGCGCAAGCGG 3942360
QY 524 GG 525
Db 3942359 GG 3942358

RESULT 6
US-08-402-282-3
Sequence 3, Application US/08402282
Patent No. 5476768
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,282
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Puglit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"

OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
OTHER INFORMATION: sequence"
OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"

Db	1078	yy	1137
Qy	187	TATCGGGCGGCACTGCTGCAGCTGCATACCGAGGGGCTCACCCTTCGAGGCGGTGTGC	246
Db	1138	yy	1197
Qy	247	ACCGGTACAACACTTCAGAGCTGGGGCGGCCCCCACTCCGGCTATGGCGGA	306
Db	1198	yy	1257
Qy	307	GTAAGTATCAGAGCCAGTGATGCGAATAAATTCCTACAGGGGTGTTACGCTGCATAC	366
Db	1258	yy	1317
Qy	367	CGGTAGCCGACCCACCCCTGCACGTGCTGCTACAGTACGATGACGAGATT	426
Db	1318	yy	1377
Qy	427	TATGCTGCGAACCCCTACCAACACACTTGCTCGAGCCCACTGACGGCGTGTGCC	486
Db	1378	yyGTA	1437
Qy	487	ATGATGCTTTGGCGCCTCGACGATGCCAACAAGTAC	525
Db	1438	CCAATTCCTCTACTTTAACTACTGATGATGATAG	1476

```

1      RESULT 4
2      US-08-323-443B-1
3      : Sequence 1, Application US/08323443B
4      : Patent No. 5654170
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: KLINGER, KATHERINE W.
9      : APPLICANT: LINDER, GREGORY M.
10     : APPLICANT: BURN, TIMOTHY C.
11     : APPLICANT: CONNORS, TIMOTHY D.
12     : APPLICANT: DACKOWSKI, WILLIAM R.
13     : APPLICANT: GERMINO, GREGORY
14     : APPLICANT: QIAN, FENG
15     :
16     : TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
17     :
18     : NUMBER OF SEQUENCES: 8
19     : CORRESPONDENCE ADDRESS:
20     :
21     : ADDRESSEE: Darby & Darby PC
22     : STREET: 805 Third Avenue
23     : CITY: New York
24     : STATE: NY
25     :
26     : COUNTRY: USA
27     :
28     : ZIP: 10022
29     :
30     : COMPUTER READABLE FORM:
31     :
32     : MEDIUM TYPE: Floppy disk
33     :
34     : COMPUTER: IBM PC compatible
35     : OPERATING SYSTEM: PC-DOS/MS-DOS
36     : SOFTWARE: PatentIn Release #1.0, Version #1.30
37     :
38     : CURRENT APPLICATION DATA:
39     :
40     : APPLICATION NUMBER: US/08/323,443B
41     : FILING DATE: 12-OCT-1994
42     : CLASSIFICATION: 435
43     :
44     : ATTORNEY/AGENT INFORMATION:
45     :
46     : NAME: Ludwig, S. Peter
47     : REGISTRATION NUMBER: 25,351
48     : REFERENCE/DOCKET NUMBER: 0372/0A462
49     : TELECOMMUNICATION INFORMATION:
50     :
51     : TELEPHONE: (212) 527-7700
52     : TELEFAX: (212) 753-6237
53     :
54     : INFORMATION FOR SEQ ID NO: 1:
55     :
56     : SEQUENCE CHARACTERISTICS:
57     :
58     : LENGTH: 31571 base pairs
59     : TYPE: nucleic acid
60     : STRANDEDNESS: single
61     : TOPOLOGY: linear
62     :
63     : MOLECULE TYPE: DNA (genomic)
64     :
65     : HYPOTHEetical: NO

```

;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; IMMEDIATE SOURCE:
;; CLONE: PKD1 GENOMIC
;;
US-08-323-443B-1

Query Match	6.0%;	Score 36.8;	DB 1;	Length 31571;
Best Local Similarity	47.3%;	Pred. No. 0.39;		
Matches 95; Conservative	0;	Mismatches 106;	Indels 0;	Gaps 0;

Qy	52	CCATTTGTTGGGGGCGGCTCTACAGCCCGGACCTTTATGACAGGACGGGTCTGTGGCCAG	111
Db	25427	CCAGGTCTGTGGTGAACACATGTATCGACACCTTACGCTGCCNNACTGGGTAGAGAGG	25486
Qy	112	GCCCAACAGAGAGGATCTTTCATGTACAGTGGCCCCCAGTTCACTTGATATATCTTGTGA	171
Db	25487	GCCNANNNTTGGGGGNGTGGACAGGAAAGTGGGCGCTGAACGTGTGCTTCCGCGCTCCCC	25546
Qy	172	ATGCCCTGGCTTTTCATATTCGGGCGGCCACTGTCTGTACACTGTGATACGAGGGGCTCACTT	231
Db	25547	GGGCCCTGGCTCTTGTGTCTGTCTGTCCGCCCGAGTGCACACTGTGATGAGGCGGTGCGTGC	25606
Qy	232	CGAGGCGCTGTGTGCACCGGT	252
Db	25607	CTGCGCAGGCGACGCTCTCAGTG	25627

RESULT 5

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A

; Patent No. 6294328

: APPLICANT: ELETSCHEM

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCING

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1008-05-24

NUMBER OF SEQ ID NOS: 2

```

; SOFTWARE: PatentIn Ver. 2.1

```

SEQ ID NO 2

LENGTH: 4403/b5
TYPE: DNA

ORGANISM: *Mycobacterium tuberculosis*

FEATURE:

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 5.8%; Score 35.6; DB 4; Length 4403765;

Best Local Similarity 43.6%; Pred. No. 8;

```
Matches 158; Conservative 0; Mismatches 204; Indels 0; Gaps 0;
```

164 CTTCTGCAATGCCCTGGCTTTCATATCCGGCCGCCCACTGCTGCAGCTGCATACCGAGGGG 223

b 3942719 CTGCCGCGGGCCCGCTGTGCCGCGCGGTCCCGCGCTGCCGCGCGTTCGCCGCGCGTCCG 3942660

224 CTCACCTTCGAGGCCGTGTCGCACCGGTACAACACCTTCAGAGCTGCGGCGCCCCAC 283

b 3942659 CCGTTGAGTTGCCCGCCGCTACCGCGGCCCGCGCTGCCCGCGCGTGC 3942600

284 CCCCAATCCGCGCTATTGCGGAGTAGTGATCAAGAGCCAGTGTATGGCAATAAATTGCC 343

[illegible]

b 3942599 CCGCCGTCGCCCTTGGCCGGGTCCGCCGCGCCGGTGCTGCGTTGCCGCCGTCGCC 3942540

344 403

[illegible]

b 3942539 CCGGTGCCGCCACGCCCCCGGTGCCGCCGCCGTTGCTGCCGTTGCTGAAGCTG 3942480

Qy	361	GCATACCGCTAACGCCACGACGCCACCCCGCACGCTGCTGCGCTACAGTGA-----	410
Db	1930	GCATACCGCTAACGCCACGACGCCACCCCGCACGCTGCTGCGCTACAGTGAATCAG	1989
Qy	411	-----CAGTTACGGACGAGTTT	427
Db	1990	TTTCGCTCTTCGTTGCGACGAGATGAAATTTCTTGTAAACCTCTGCGAGTTTACGAGCAGATTT	2049
Qy	428	ATGCTGCGCAGACCCCTTACACCAACACACTTGTGCTCCAGCCCCCACTTACGGCGTTGGTGCA	487
Db	2050	ATGCTGCGCAGACCCCTTACACCAACAGGCACTTGTGCTCCAGCCCCCACTTACGGCGTTGGTGCA	2109
Qy	488	TGAATGCTTTTGGCGCCTTGAACCGATGCCAAGACTAGAGACCAATGCTGATGATGGGCT	547
Db	2110	TGAATGCTTTTGGCACCCTTGTGACTGATGCCAAGACTAGAGACCAATGCTGATGATGGGCT	2169
Qy	548	TGCTTTCTTTCTTCAATTGACAGGCTATGATATACCAAGGGGATATCAACCGTTTGTCTCCAT	607
Db	2170	TGCTTTCTTTCTTCAATTGACAGGCTATGATATACCAAGGGGATATCAACCGTTTGTCTCCAT	2229
Qy	608	A 608	
Db	2230	A 2230	

```

RESULT 2
US-09-176-657-5
; Sequence 5, Application US/09176657
; Patent No. 6020164
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lu, Aina
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
; FILE REFERENCE: PF-0611 US
; CURRENT APPLICATION NUMBER: US/09/176,657
; CURRENT FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1250374
US-09-176-657-5

Query Match          10.4%; Score 63.2; DB 3; Length 1506;
Best Local Similarity 54.5%; Pred. No. 7, 2e-10;
Matches 279; Conservative 0; Mismatches 193; Indels 40; Gaps

QY      1 ATGACTAATATAAAAAGGCCGTGAACCCCTACCAATGCGTGGAAATTAATCCAGTTGTG 60
        |||
DB      768 ATGACCAATTAAGAAAGATGTGCACACCATATGCAAAATGGTTGAATTAAGCCAGTAGTT 827
        |||

QY      61 GGCGCGGTATACAGCCGCCGACTTCTATGACGAGCAGCGTGCTGT-----TGTGCCAG 111
        |||
DB      828 GGAGCGTGTATATGTGTCCGGAGTTATATGACAGCATCCAGCTTTCAAGACAGATGTGTCCCTTA 887
        |||

QY      112 GCCAACCCAGAGGAGATCTTCATGTACATGAGTGGCCCACTTCATTGTATTACTTCTGCA 171
        |||
DB      888 GGCMAATGATGCACACAGTGCCTTCATAGGAAGAGGGGGTATCAACACTTAATTCCTTTA 947
        |||

QY      172 ---ATGCTGGCTTTTGCATA---TCCGGCGCGCAGCTGCTCAGCTGCATCCGAGGGGCT 225
        |||
DB      948 ATCATTTCTGTGCTTCCCTTAACCTTACGTGCAGCACCACGSCACCCGCTTTCAAGAGAGCC 1007
        |||

QY      226 CACCTTGAAGGCGGTGTGCGACCGTGTACACACCTTCAGAGCTGTGGCGGCCCCCAACC 285
        |||

```

Db	1008	CAATTGAGGGGACAGAGGGGCGGACAGATATATGGTCAGTCCG---GGGTAACCTCCAAACA	1064
Qy	286	CCAAATCCCGGCTATATGCGCGAGTAGTAGTGTATCAAGAGCCACAGTGTATGGCAATAAATTGCTA	345
Db	1065	GGCAATCCCCCGCTATATCCAGGGGCTGGATATAGCAGCCATACAGATATGCACAGCCTGTACTTG	1124
Qy	346	CAGAGTGGTATACGCTGCATACCGCTACGGCCAGCCACCACCCCTGACCACTGCTCTCTAC	405
Db	1125	CACAC-----CGAGCCACCGCTGCTGTGCAGCCGCTGTGCAGCCGCTTAC	1165
Qy	406	AGTGCAGTTACGACACAGATTATATGCTGCCAGACCCCTTACACACACAACTTGTCTCCACC	465
Db	1166	AGTAGCGTTATGCGACAGGCTGTACACAGCCGACCCCTT---ACCAATGCCCTTGCCTCCCTCC	1222
Qy	466	CCCACTACGCGCGTTGGTGGCAGTAATGCCTT	497
Db	1223	GCTAGCTATGGAGTTGGCGCTGTGGCGAAGTTT	1254

```

127 TCTTCATCATAGTGGCCCGCATTCATTGTATATACCTTGTGCAGATGCTGCTTCCCA 186
Query Match      6.1%; Score 37.4; DB 1; Length 7218;
Best Local Similarity 5.0%; Pred. No. 0.13;
Matches 20; Conservative 204; Mismatches 175; Indels 0; Gaps 0

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:42:03 ; Search time 34.6839 Seconds
(without alignments)
5384.802 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143
Perfect score: 609
Sequence: 1 atgacataataaaggccgtt.....acaacgtcttcctccatcat 609

Scoring table: IDENTITY NUC
Gapop 10'-0', Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428.6	70.4	2372	4 US-09-145-391-1	Sequence 1, Appli
2	63.2	10.4	1506	3 US-09-176-657-5	Sequence 5, Appli
3	37.4	6.1	7218	1 US-08-232-463-14	Sequence 14, Appli
4	36.8	6.0	31571	1 US-08-323-443B-1	Sequence 1, Appli
5	35.6	5.8	4403765	4 US-09-103-840A-2	Sequence 2, Appli
6	34.2	5.6	15664	1 US-08-402-282-3	Sequence 3, Appli
7	34.2	5.6	15664	1 US-08-508-004-3	Sequence 3, Appli
8	34.2	5.6	15664	1 US-08-402-066-3	Sequence 3, Appli
9	34.2	5.6	15664	1 US-08-402-068-3	Sequence 3, Appli
10	33.4	5.5	538	4 US-09-056-556-180	Sequence 180, App
11	33.4	5.5	538	4 US-09-072-596-175	Sequence 175, App
12	32.8	5.4	535	4 US-09-056-556-171	Sequence 171, App
13	32.8	5.4	535	4 US-09-072-596-166	Sequence 166, App
14	32.2	5.3	1071	2 US-08-997-080-180	Sequence 180, App
15	32.2	5.3	1071	2 US-08-997-362-180	Sequence 180, App
16	32.2	5.3	1071	4 US-09-095-855-180	Sequence 180, App
17	32.2	5.3	1071	4 US-09-324-542-180	Sequence 180, App
18	32.2	5.3	1071	4 US-09-205-426-180	Sequence 180, App
19	31.6	5.2	610	4 US-09-072-596-293	Sequence 293, App
20	31.6	5.2	418	3 US-09-141-000-6	Sequence 6, Appli
21	31.4	5.2	458	3 US-09-141-000-4	Sequence 4, Appli
22	31.4	5.2	441529	4 US-09-103-840A-1	Sequence 1, Appli
23	31.2	5.1	2558	3 US-08-999-733-3	Sequence 3, Appli
24	31.2	5.1	9757	1 US-08-093-453B-1	Sequence 1, Appli
25	31.2	5.1	9759	1 US-08-453-041A-1	Sequence 1, Appli
26	31.2	5.1	9759	3 US-08-999-733-1	Sequence 1, Appli
27	30.8	5.1	441529	4 US-09-103-840A-1	Sequence 1, Appli

C	28	30.6	5.0	4793	4 US-09-561-497-10	Sequence 10, Appli
	29	30	4.9	3639	4 US-09-605-785-779	Sequence 779, App
	30	30	4.9	3848	4 US-09-112-096-28	Sequence 28, Appli
	31	30	4.9	5668	4 US-09-112-096-14	Sequence 14, Appli
	32	30	4.9	5668	4 US-09-605-785-777	Sequence 777, App
	33	29.8	4.9	744	4 US-09-385-028-22	Sequence 22, Appli
	34	29.8	4.9	2283	4 US-09-153-804-5	Sequence 5, Appli
	35	29.8	4.9	2903	2 US-08-310-912A-1	Sequence 1, Appli
	36	29.8	4.9	2903	3 US-08-841-089-1	Sequence 1, Appli
	37	29.8	4.9	2903	4 US-09-301-085-1	Sequence 1, Appli
	38	29.8	4.9	2903	5 PCT-US95-04570-1	Sequence 1, Appli
	39	29.8	4.9	2903	5 PCT-US95-04589-1	Sequence 1, Appli
	40	29.8	4.9	3432	2 US-08-310-912A-141	Sequence 141, App
	41	29.8	4.9	3432	4 US-09-301-085-141	Sequence 141, App
	42	29.8	4.9	3432	5 PCT-US95-04589-141	Sequence 141, App
	43	29.8	4.9	11604	4 US-09-385-028-13	Sequence 13, Appli
	44	29.8	4.9	15079	4 US-09-385-028-1	Sequence 1, Appli
	45	29.6	4.9	11673	4 US-09-334-220-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-145-391-1
; Sequence 1, Application US/09145391
; Patent No. 6194171
; GENERAL INFORMATION:
; APPLICANT: Pulat, Stefan M.
; APPLICANT: Shiba, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/145.391
; CURRENT FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987)..(1979)
US-09-145-391-1

Query	Match	Score	70.4%	DB 4:	Length 2372;
Best Local Similarity	82.0%	Pred. No. 2.6e-122;		Indels	55; Gaps 2;
Matches	542;	Conservative	0;	Mismatches	64;
QY	1	ATGACTAATATAAAGCCGTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG	60		
DB	1572	ATGACAAATATAAAGACCCGTCAACCCCTATACAAATGCTGGAATTAATCCAGTTGTG	1631		
QY	61	GGCGGCTTCAAGCCCGGACTTTCAGAGGACGAGTGTGTTGTCAGAGCCACAG	120		
DB	1632	GATGAGCTTCAAGCCCGGACTTTCAGAGGACGAGTGTGTTGTCAGAGCCACAG	1691		
QY	121	GAGGATTTTCAGTATAGAGGCCCCAGTTCATTGATATATCTTCGCAATGCTGGC	180		
DB	1692	GAGGATTTTCAGTATAGAGGCCCCAGTTCATTGATATATCTTCGCAATGCTGGC	1751		
QY	181	TTTCATATCCGCGCGGCTGCTGAGCTGCATACGAGGGGCTCACCTTCGAGGCGGT	240		
DB	1752	TTTCATATCCGCGCGGCTGCTGAGCTGCATACGAGGGGCTCACCTTCGAGGCGGT	1811		
QY	241	GGTGCACCGTGTACCAACACCTTCAGAGCTCGGGCCCAACCCCAATCCGGCTAT	300		
DB	1812	GGTGCACCGTGTACCAACACCTTCAGAGCTCGGGCCCAACCCCAATCCGGCTAT	1871		
QY	301	GGCGAGTGTGTATCAAGAGCCAGTGTATGCAATTAATTCACAGGGTGTACCT	360		
DB	1872	GGCGAGTGTGTATCAAGAGCCAGTGTATGCAATTAATTCACAGGGTGTACCT	1929		

QY 252 GTACAAACCTTCAGAGCTGGGGGCCCCACCCCAATCCCGGCTATGGCGAGTAGT 311
 Db 1 GTACAAACCTTCAGAGCTGGGGGCCCCCGCCCGGCTATCCCGGCTACGGCGAGTGT 60
 QY 312 GTATCAAGAGCAGTGTATGGCAATTAATGTCTACAGGGTGGTTCAGCTGATACCGCTA 371
 Db 61 TTACCAAGATGATTTTATGCG--TGACAGACTTTATGTGTGTATGCTGATACCGCTA 117
 QY 372 CGCCAGCCCACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 431
 Db 118 CGCCAGCTACCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177
 QY 432 TGGCAGCCCTTACCAACACACTTGTCTCAGCCCCCACTTACGGCGTGGTGGCTATGA 491
 Db 178 TGGCAGCCCTTACCAACACTTGTCTCAGCCCCCACTTACGGCGTGGTGGCTATGA 237
 QY 492 TGGCTTTGGCGCTTACCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 551
 Db 238 TGGCTTTGGCGCTTACCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 297
 QY 552 TCTTTCTTCATGAGGCTAGTATATACCAAGGGGATACACCGTTTGGCTCCATA 608
 Db 298 TCTTTCTTCATGAGGCTAGTATATACCAAGGGGATACACCGTTTGGCTCCATA 354

RESULT 14
 A1968176 439 bp mRNA linear EST 25-AUG-1999
 LOCUS w11406.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2516914.3
 DEFINITION similar to TR:043251 043251 HYPOTHETICAL 39.5 KD PROTEIN. ;, mRNA
 sequence.
 ACCESSION A1968176
 VERSION A1968176.1 GI:5764994
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 439)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.lnl.gov/bdip/image/image.html
 Seq primer: -40UP from G1bco.
 Location/Qualifiers

FEATURES

source
 1. 439
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2516914"
 /clone_1b="NCI CGAP GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
 from the normalized library NCI CGAP GC6 was prepared, and
 as circles were made in vitro. Following HAP hybridization,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clones
 1257096-1258631, 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldo."
 91 a 142 c 114 g 92 t

BASE COUNT

ORIGIN

Query Match 45.6%; Score 277.6; DB 9; Length 439;
 Best Local Similarity 84.7%; Pred. No. 2.3e-71;
 Matches 342; Conservative 0; Mismatches 19; Indels 43; Gaps 1;

QY 205 GCAGCTGATTCAGAGGGGCTCACCTTCAGAGCGGTGTGTCACCGGTATCAACACTTC 264
 Db 1 GCGGCGGCTTACAGAGGGGCGACCTCGAGGCGCGGTGTGTCACCGGTATCAACACTTC 60
 QY 265 AGAGCTGGGGGCGCCCGCCCAATCCCGGCTATGCGGAGAGTATCAAGAGCA 324
 Db 61 AGGCGCGGGGCGCCCGCCCGCCCGGCTATGCGGAGAGTATCAAGAGCA 100
 QY 325 GTGTATGCAATTAATTGCTACAGGTGTGTTACGCTGCATACCGCTACGCCACCGC 384
 Db 101 -----GGGTGTATGCTGCATACCGCTACGCCACCGC 137
 QY 385 CTTGCCACTGCTGCTGCTTACAGTACAGTACAGAGTATGCTGCGGACCTTAC 444
 Db 138 CTTGCCACTGCTGCTGCTTACAGTACAGTACAGAGTATGCTGCGGACCTTAC 197
 QY 445 CACACACACTGCTGCTGCGAGCCCGCCAGCTAGCGGCTGTCATGATGCTTTGGCGCC 504
 Db 198 CACACACACTGCTGCTGCGAGCCCGCCAGCTAGCGGCTGTCATGATGCTTTGGCACT 257
 QY 505 TTGACGATGCGCAAGACTAGAGGCGCATGATGATGATGATGATGATGATGATGATGAT 564
 Db 258 TTGACGATGCGCAAGACTAGAGGCGCATGATGATGATGATGATGATGATGATGATGAT 317
 QY 565 CAGGCTGATATATCCAAAGGGGATACACCGTTTGGCTCCATA 608
 Db 318 CAGGCTGATATATCCAAAGGGGATACACCGTTTGGCTCCATA 361

RESULT 15

AM589795 439 bp mRNA linear EST 22-MAR-2000
 LOCUS hg22d04.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2946343.3
 DEFINITION similar to TR:043251 043251 HYPOTHETICAL 39.5 KD PROTEIN. ;, mRNA
 sequence.
 ACCESSION AM589795
 VERSION AM589795.1 GI:7276916
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 439)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 image.lnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from G1bco.
 Location/Qualifiers

FEATURES

source
 1. 439
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2946343"

FEATURES

source
 1. 439
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2946343"
 Possible reversed clone: similarity on wrong strand
 Possible reversed clone: polyT not found
 Seq primer: -40UP from G1bco.

```
QY 523 AGAGCCATGCTGATGATGCTGCTGCTTTCTTTCTTCAATTCAGGCTAGTATATACCA 582
DB 298 AGAGCCATGCTGATGATGCTGCTGCTTTCTTTCTTCAATTCAGGCTAGTATATACCA 357
QY 583 GGGGATACACCGCTTTGCTCCATA 608
DB 358 GGGGATACACCGCTTTGCTCCATA 383

RESULT 12
AUI45000/c 607 bp mRNA linear EST 05-AUG-2002
LOCUS AUI45000 HEMBA1 Homo sapiens cDNA clone HEMBA1003591 3', mRNA
DEFINITION sequence.
ACCESSION AUI45000
VERSION AUI45000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 607)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saio,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
,S. and Isogai,T.).
HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
Saio,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
,Y., Nagai,T., Sugano,S., Isogai,T.)
Unpublished (2000)
JOURNAL Contact: Takao Isogai
COMMENT Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source location/Qualifiers
1..607
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003591"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME18SFL3"
BASE COUNT 140 a 152 c 179 g 129 t 7 others
ORIGIN
Query Match 49.2%; Score 299.8; DB 9; Length 607;
Best Local Similarity 79.1%; Pred. No. 7e-78;
Matches 389; Conservative 0; Mismatches 99; Indels 4; Gaps 3;

QY 121 GAGGATCTTCATGATGCTG-GCCCGATTCATTGATATCTTCAGATGCTGG 179
DB 595 GAGATTTTTCATGACACAGCAGCCCGCCCGATTCATTGATATCTTCAGATGCTGG 536
QY 180 CTTTCC-ATATCGGCGCGGCTGCTGATCGATACCGAGGCGCTCACTTGAAGC 237
DB 535 GCTTTCGGATATCCAGAGCCACCGCGCGCGCTTACCGAGGCGCGCTTGCAGAGC 476
QY 238 CG-TGCTGCAACCGTTCACACCTTCAGAGCTGGCGCGCGCCCGCCCAATCCCGC 256
DB 475 CGGCGGTGACCGTTCACACCTTCAGAGCGCGCGCGCGCGCGCCCGCCCAATCCCG 416
QY 297 CTATGCGGAGTATGATATCAAGAGCAGTATGCAATTAATTCAGAGGCTGTTA 356
DB 415 GCCTAGCGCGGCTTCTTACAGATGATTTTATGTCAGACATTTATGCTGTTA 356
QY 357 CGGTGATACCGTACGCGCCAGCCCGCTGCATGCTGCTTACAGTACAGTTA 416
```

```
DB 355 TCGTCATACCGCTACGCGCCAGCCCTACCCCTGCGCATGCTACAGTACAGTTA 296
QY 417 CGAGCATTTATGCTGCGAGCCCTACACACACACTTCTCAGGCCCCACCTACAG 476
DB 295 CGAGCATTTATGCTGCGAGCCCTACACACACACTTCTCAGGCCCCACCTACAG 236
QY 477 CGTTGTCATGAAATGCTTTTGGCGCCCTTACCGATCCAAAGACTAGAGCATGCTGA 536
DB 235 CGTTGTCATGAAATGCTTTTGGCGCCCTTACCGATCCAAAGACTAGAGCATGCTGA 176
QY 537 TGATGTGGTCTCGTTCTTTCTTCAATTCAGGCTAGTATATACCAAGGGGATACACG 596
DB 175 TGATGTGGTCTCGTTCTTTCTTCAATTCAGGCTAGTATATACCAAGGGGATACACG 116
QY 597 TTTTGTCTCCATA 608
DB 115 TTTTGTCTCCATA 104

RESULT 13
A1244212 459 bp mRNA linear EST 28-JAN-1999
LOCUS A1244212
DEFINITION q186b11.x1 NCI_CGAP_kid3 Homo sapiens cDNA clone IMAGE:186633 3',
mRNA sequence.
ACCESSION A1244212
VERSION A1244212
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 459)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgap@nci.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNI, at:
www.bio.lnl.gov/bbrp/image/image.html
Insert length: 452 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 451.
FEATURES
source location/Qualifiers
1..459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:186633"
/clone_lib="NCI CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidney. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
BASE COUNT 105 a 132 c 108 g 112 t 2 others
ORIGIN
Query Match 47.6%; Score 289.8; DB 9; Length 459;
Best Local Similarity 90.2%; Pred. No. 5.6e-75;
Matches 322; Conservative 0; Mismatches 32; Indels 3; Gaps 1;
```

ACCESSION A1374891
VERSION A1374891.1 GI:4174881
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 400 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
Source
1..449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2048518"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCCTTAATTTTCTTTTCTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 93 a 136 c 114 g 105 t 1 others
ORIGIN
Query Match 51.6%; Score 314; DB 9; Length 449;
Best Local Similarity 90.2%; Pred. No. 3,6e-82;
Matches 348; Conservative 0; Mismatches 35; Indels 3; Gaps 1;
QY 223 GCTACCTTGAGGCGGTGTCGACCGGTGACCAACCTTCAGAGCTGCGCGCCCA 282
Db 1 GCGCACCCTGCGAGGCGCGGTGTCGACCGGTGACCAACCTTCAGAGGCGCGCCCG 60
QY 283 CCCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGACCGAGTATGCGCAATTAATG 342
Db 61 CCCCCGATCCCGGCTATGCGGAGTAGTGTATCAAGAGTATTAATG--TGCAGAC 117
QY 343 CTACAGGCTGTTACGTCATACCGCTACGCGCGACCCCGCCACCTGCTGCGC 402
Db 118 ATTTATGCTGTTATGCTGATACCGCTACCGCGCCCTACCGCTGCGCGCTGCGC 177
QY 403 TACAGTACAGTTACGAGCAGATTATGCTGCGACCCCTACCAACAGACACTTGTCTCA 462
Db 178 TACAGTACAGTTACGAGCAGATTATGCTGCGACCCCTACCAACAGACACTTGTCTCA 237
QY 463 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGGCGCCCTTACCGAGTCCAAAGCT 522
Db 238 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGGACCTTTTACATGATGCCAAAGCT 297
QY 523 AGAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
Db 298 AGAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
QY 583 GGGGGATACAAACCTTTTGTCTCATTA 608
Db 358 GGGGGATACAAACCTTTTGTCTCATTA 383
RESULT 11

A1917298
LOCUS A1917298 456 bp mRNA linear EST 14-DEC-1999
DEFINITION ts96g09.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2239168 3',
RNA sequence.
ACCESSION A1917298
VERSION A1917298.1 GI:5637153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 456)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskalko, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/<http://www-bio.llnl.gov/bdrip/image/image.html>
Insert length: 478 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
Source
1..456
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2239168"
/clone_lib="NCI CGAP GC6"
/issue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
as circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 97 a 137 c 115 g 107 t
ORIGIN
Query Match 51.6%; Score 314; DB 9; Length 456;
Best Local Similarity 90.2%; Pred. No. 3,7e-82;
Matches 348; Conservative 0; Mismatches 35; Indels 3; Gaps 1;
QY 223 GCTACCTTGAGGCGGTGTCGACCGGTGACCAACCTTCAGAGCTGCGCGCCCA 282
Db 1 GCGCACCCTGCGAGGCGCGGTGTCGACCGGTGACCAACCTTCAGAGGCGCGCCCG 60
QY 283 CCCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGACCGAGTATGCGCAATTAATG 342
Db 61 CCCCCGATCCCGGCTATGCGGAGTAGTGTATCAAGAGTATTAATG--TGCAGAC 117
QY 343 CTACAGGCTGTTACGTCATACCGCTACGCGCGACCCCGCCACCTGCTGCGC 402
Db 118 ATTTATGCTGTTATGCTGATACCGCTACCGCGCCCTACCGCTGCGCGCTGCGC 177
QY 403 TACAGTACAGTTACGAGCAGATTATGCTGCGACCCCTACCAACAGACACTTGTCTCA 462
Db 178 TACAGTACAGTTACGAGCAGATTATGCTGCGACCCCTACCAACAGACACTTGTCTCA 237
QY 463 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGGCGCCCTTACCGAGTCCAAAGCT 522
Db 238 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGGACCTTTTACATGATGCCAAAGCT 297

Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/URL at:
www-bio.lnlnl.gov/bbrip/image/image.html
Insert Length: 460 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES
source

1. 480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2119486"
/clone_1id="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI CGAP Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 101 a 146 c 124 g 108 t 1 others
ORIGIN

Query Match 53.7%; Score 327.2; DB 9; Length 480;
Best Local Similarity 89.9%; Pred. No. 4.6e-86;
Matches 363; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

QY 205 GCAGCTGCATACCGAGGGGCTCACTTCGAGGCGCGTGTGCGACCGCTGACAAACACTTTC 264
DB 1 GGGGCGGCTTACCGAGGGGCGACCTCGAGGCGCGGTGCGACCGCTGACAAACACTTTC 60
QY 265 AAGAGTGGCGGGCGGGCCCAATCCGGGCTTATGCGAGAGTAGTATGATCAGAGCCA 324
DB 61 AAGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 120
QY 325 GTGTATGGCAATTAATTTGTAAGGGTGTGTAAGCTGATACCGCTACGCCCA3CCCA 384
DB 121 TTTTATGG---TGACAGACATTTATGTGTATGTGATGACATACCGCTACGCCCA3CC 177
QY 385 CCTGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 444
DB 178 CCTGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 237
QY 445 CAGCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 504
DB 238 CAGCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297
QY 505 TTGACCGATGCCAAGACTAGAGCCATGCTGATGATGTGGTCTGTTCTTTCTTCATTG 564
DB 298 TTGACCGATGCCAAGACTAGAGCCATGCTGATGATGTGGTCTGTTCTTTCTTCATTG 357
QY 565 CAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCATA 608
DB 358 CAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCATA 401

RESULT 9
A1189407 383 bp mRNA linear EST 28-OCT-1998
LOCUS A1189407
DEFINITION qd01h05.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1722489 3', mRNA sequence.
ACCESSION A1189407
VERSION A1189407.1 GI:3740616
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnlnl.gov) for further information.
Insert Length: 486 Std Error: 0.00
Seq primer: -40UP from Gibco.
High quality sequence stop: 381.
Location/Qualifiers
1. 383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1722489"
/clone_1id="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer [5',
AAGTGAAGAATTCG
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT
ORIGIN

Query Match 51.6%; Score 314; DB 9; Length 383;
Best Local Similarity 90.2%; Pred. No. 3.4e-82;
Matches 348; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

QY 214 TACGAGGGGCTCACCTTCGAGGCGCGTGTGCGACCGCTGACAAACACTTTCAGAGCTGCG 273
DB 1 TACGAGGGGCTCACCTTCGAGGCGCGTGTGCGACCGCTGACAAACACTTTCAGAGCTGCG 60
QY 274 GCGCCCCCAACCCCAATCCCGGCTATGCGAGAGTAGTATGATCAGAGCCAGTATGCG 333
DB 61 GCGCCCCCGCCCGCATCCGGGCTTACGAGGCGGGTGTGTTACAGAGATGATTTATG9- 119
QY 334 AATTAATGCTACAGGTTGTTACGCTGATACCGCTACGCCCAACCCCTGCGACT 393
DB 120 --TGACAGACATTTATGTTATGTTATGCTGATACCGCTACGCCCAACCCCTGCGACT 177
QY 394 GGTGCTGCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 453
DB 178 GCGGCTGCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 237
QY 454 CTGTGCTCAGCCCCCAACCTTACGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513
DB 238 CTGTGCTCAGCCCCCAACCTTACGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
QY 514 GCCAAGACTAGAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 573
DB 298 GCCAAGACTAGAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 357
QY 574 ATATACCAAGGGGATACCAACCGTTT 599
DB 358 ATATACCAAGGGGATACCAACCGTTT 383

RESULT 10
A1374891 449 bp mRNA linear EST 16-FEB-1999
LOCUS A1374891
DEFINITION ta60g12.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:2048518 3', mRNA sequence.

BASE COUNT 184 a 232 c 185 g 170 t 2 others
ORIGIN
Query Match 55.6%; Score 338.8; DB 10; Length 773;
Best Local Similarity 88.4%; Pred. No. 2,1e-89;
Matches 380; Conservative 0; Mismatches 47; Indels 3; Gaps 1;
Qy 179 GCTTTCATATCCGGCCGACCTGCTGAGTGCATACCGAGGGGCTCACTTCGAGGCC 238
Db 1 GCTTCCCGATACAGCAGCCACCGTCGCGCGCTTACCGAGGGGCGACCTCGAGGCC 60
Qy 239 GTGTCGACCGGTGACAAACACCTTCAGAGTCGCGCGCCCAACCCCAATCCCGGCT 298
Db 61 GCGGTGCGACCGGTGACAAACCTTCAGAGCGCGCGCCCGCCCGCATCCGCGCT 120
Qy 299 ATGCGGAGTATGATGATGAAGAGCCAGTGTATGGATAAATGCTACAGGGTGTACG 358
Db 121 ACGGCGGTGTGTTTACAGATGATGATTTATG---TGACAGCATTTATGTTGTTATG 177
Qy 359 CTGCATACCGGTACGCGCCGACCCCTGCGCATGCTGCTGCTACAGTACAGTACG 418
Db 178 CTGCATACCGGTACGCGCCGACCTGCGCATGCTGCTGCTACAGTACAGTACG 237
Qy 419 GACGAGTTATGCTGCGCAGCCCTTACACACACACTTGTCTCAGACCCCACTTACGGCG 478
Db 238 GACGAGTTATGCTGCGCAGCCCTTACACACACACTTGTCTCAGACCCCACTTACGGCG 297
Qy 479 TTGGTGCATATGCTTTTGGCCCTTGCAGATGAGCCAGACTGAGAGCCATGCTGATG 538
Db 298 TTGGTGCATATGCTTTTGGCCCTTGCAGATGAGCCAGACTGAGAGCCATGCTGATG 357
Qy 539 ATGTGGTCTGCTTCTTCTTCTTATGACGAGCTGATATATACCAAGGGGATACAACTT 598
Db 358 ATGTGGTCTGCTTCTTCTTCTTATGACGAGCTGATATATACCAAGGGGATACAACTT 417
Qy 599 TTGCTCCATA 608
Db 418 TTGCTCCATA 427
RESULT 7
AM197589 477 bp mRNA linear EST 29-NOV-1999
LOCUS xM44902.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2687090 3'
DEFINITION similar to TR:Q43251 Q43251 HYPOTHETICAL 39.5 KD PROTEIN.; mRNA
sequence.
ACCESSION AM197589
VERSION AM197589.1 GI:6476819
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Seq primer: -40UP from Gidco.

FEATURES
source
Location/Qualifiers
1. .477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2687090"
/clone_1lb="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257086-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 100 a 145 c 123 g 109 t
ORIGIN
Query Match 53.7%; Score 327.2; DB 10; Length 477;
Best Local Similarity 89.9%; Pred. No. 4.6e-86;
Matches 363; Conservative 0; Mismatches 38; Indels 3; Gaps 1;
Qy 205 GCACTGCATACCGAGGGGCTACCTTGAGGCGGTGCTGCGCACCGGTATACACACTTC 264
Db 1 GCGGCGGCTTACCGAGGGGCGCACCTGCGAGCGCGGTGCGCACCGGTATACACACTTC 60
Qy 265 AGAGCTGGGGGCGCCCAACCCCAATCCCGGCTATGGCGGAGTATACAGAGCCA 324
Db 61 AGGCGCGGGGCGCCCGCCCGATCCGGCTACGCGGAGTGTGTTACAGAGTGA 120
Qy 325 GTGTATGCAATAAATTGCTACAGAGGTGTTACCGCTGATACCGCCAGCCCACT 384
Db 121 TTTTATG---TGACAGATTATGATGTTATGCTGATACCGCTACCGCCAGCTAC 177
Qy 385 CTGCGCATGCTGCTGCTTACAGTGAAGTACGAGAGATTATGCTGCGGACCTTAC 444
Db 178 CTTGCGCATGCTGCTGCTTACAGTGAAGTACGAGAGATTATGCTGCGGACCTTAC 237
Qy 445 CACCAACACTTGTCCAGACCCCACTAGCGCGTGTGTCATGAATGCTTTGCGGCC 504
Db 238 CACCAACACTTGTCCAGACCCCACTAGCGCGTGTGTCATGAATGCTTTGCGACCT 297
Qy 505 TTGACCATGCGCAAGATGAGAGCCATGATGATGAGTCTGCTTCTTCTTCAATTG 564
Db 298 TTGACATGTCGAAGATGAGAGCCATGATGATGAGTCTGCTTCTTCTTCAATTG 357
Qy 565 CAGGCTAGTATATACCAAGGGGATACAACTTTTCTCCATA 608
Db 358 CAGGCTAGTATATACCAAGGGGATACAACTTTTCTCCATA 401
RESULT 8
AI401040 480 bp mRNA linear EST 30-MAR-1999
LOCUS th27a12.x1 NCI CGAP P-28 Homo sapiens cDNA clone IMAGE:2119486 3',
DEFINITION mRNA sequence.
ACCESSION AI401040
VERSION AI401040.1 GI:4244127
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 480)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Best Local Similarity 94.6%; Pred. No. 4.5e-95;
Matches 370; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 205 GCAGTCGATACCGAGGGGCTGACCTTGCAGGCGCGTGTGCGACCGGTGACAAACCTTC 264
Db 1 GCGGCGGCTTACCGAGGGGCGACCTTGCAGGCGCGGTGCGACCGGTGACAAACCTTC 60
Qy 265 AGAGTCGGGCGCGCCCAACCCCAATCCCGGCTATGGCGGAGTAGTATGACAGGCCA 324
Db 61 AGGGCGGGGGCGCGCCCGCGCGATCCGGCTACGGCGGAGTAGTATGACAGGCCCT 120
Qy 325 GTGTATGCGAATAATTTGTTACAGGGTGTACGCTGCTACCCGCTACCCGACCCAC 384
Db 121 GTGTATGCGAATAATTTGTTACAGGGTGTATGCTGCTACCCGCTACCCGACCCAC 180
Qy 385 CCTGCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 444
Db 181 CCTGCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Qy 445 CACCAACACCTTCTCTCAAGCCCGCCCACTACGGCGTGTGCTGCTGCTGCTGCTG 504
Db 241 CACCAACACCTTCTCTCAAGCCCGCCCACTACGGCGTGTGCTGCTGCTGCTGCTGCTG 300
Qy 505 TTGACCGATGCCAAGACTTGAGGCCATGCTGATGATGCTGCTGCTGCTGCTGCTGCT 564
Db 301 TTGACCGATGCCAAGACTTGAGGCCATGCTGATGATGCTGCTGCTGCTGCTGCTGCT 360
Qy 565 CAGGCTAGTATATACCAAGGGGATATCAAC 595
Db 361 CAGGCTAGTATATACCAAGGGGATATCAAC 391

RESULT 5

AV729057 774 bp mRNA linear EST 17-OCT-2000
LOCUS AV729057 HTC Homo sapiens cDNA clone HTCB8601 5', mRNA sequence.
DEFINITION AV729057
ACCESSION AV729057
VERSION AV729057.1 GI:10838478
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS

Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, X., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu,
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
Contact: Zeguana Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source

1..774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCB8601"
/clone_lib="HTC"
/issue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT

186 a 217 c 180 g 191 t

Query Match

56.2%; Score 342; DB 10; Length 774;

Best Local Similarity 88.8%; Pred. No. 2.4e-90;
Matches 382; Conservative 0; Mismatches 45; Indels 3; Gaps 1;

Qy 179 GCTTCCATATCCGGCCCGCCACCTGCTGAGCTGCAATACCGAGGGGCTCACCTTGAGGCC 238
Db 1 GCTTCCATATCCGGCCCGCCACCTGCTGAGCTGCAATACCGAGGGGCTCACCTTGAGGCC 60
Qy 239 GTGTGCGACCGGTATGACCAACCTTACAGAGTGGCGCGCCCGCCCAATCCCGGCT 298
Db 61 GCGGTGCGACCGGTATGACCAACCTTACAGAGTGGCGCGCCCGCCCAATCCCGGCT 120
Qy 299 ATGGCGGAGTAGTATGATCAAGAGCAGTGTATGGCAATAATTTGTAAGGTTGATAG 358
Db 121 ACGGGGGGTGTATGATCAAGAGTGTATGGCAATAATTTGTAAGGTTGATAG 177
Qy 359 CTGCATACCGCTACCGCCAGCCAGCCAGCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCT 418
Db 178 CTGCATACCGCTACCGCCAGCCAGCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 237
Qy 419 GACGAGTTATGCTGCGCGACCCCTTACCAACACACTTGTCCAGCCCGCCACTACGGCG 478
Db 238 GACGAGTTATGCTGCGCGACCCCTTACCAACACACTTGTCCAGCCCGCCACTACGGCG 297
Qy 479 TTGGTCCCATGATGCTTTTGGCGCCCTTGACCGAGTCCCAAGACTGAGCATGCTGATG 538
Db 298 TTGGTCCCATGATGCTTTTGGCGCCCTTGACCGAGTCCCAAGACTGAGCATGCTGATG 357
Qy 539 ATGTGGGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 598
Db 358 ATGTGGGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 417
Qy 599 TTGCTCCATA 608
Db 418 TTGCTCCATA 427

RESULT 6

AV729198 773 bp mRNA linear EST 17-OCT-2000
LOCUS AV729198 HTC Homo sapiens cDNA clone HTCB8601 5', mRNA sequence.
DEFINITION AV729198
ACCESSION AV729198
VERSION AV729198.1 GI:10838619
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS

Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, X., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu,
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
Contact: Zeguana Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source

1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCB8601"
/clone_lib="HTC"
/issue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

```

|||||
Db 236 CTACAGTACAGTTAGCGAGCGATTATGCTGCCAGCCCTTACACACACACCTTGCTCC 177
Oy 462 AGCCCCCACTACGCGCTTGCTGTCATGAATGCTTTTCGCGCCCTTGACCGATGCCAGAC 521
Db 176 AGCCCCCACTACGCGCGCTTGCTGTCATGAATGCTTTTCGCGCCCTTGACCGATGCCAGAC 117
Oy 522 TAGAGACCATCTGATGATGTGGGTCTCGTTCTTTCTTCTTATGACGCTAGTATATACCA 581
Db 116 TAGAGACCATCTGATGATGTGGGTCTCGTTCTTTCTTCTTATGACGCTAGTATATACCG 57
Oy 582 AGGGGATACACCGTTTGTCCCATAT 609
Db 56 AGGGGATACACCGTTTGTCCCATAT 29

RESULT 3
LOCUS A1656926 495 bp mRNA linear EST 04-MAY-1999
DEFINITION U48808.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2244038 3'
similar to TR:Q43251 Q43251 HYPOTHETICAL 39.5 KD PROTEIN. ; mRNA
sequence.
A1656926
A1656926.1 GI:4740905
EST.
KEYWORDS
ACCESSION A1656926.1 GI:4740905
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 495)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Straubeberg, Ph.D.
Email: rgs@bbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmerich-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnlnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 487.
Location/Qualifiers
FEATURES
SOURCE
1. .495
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2244038"
/clone_lib="NCI-CCAP GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI-CCAP GC4 was prepared, and
88 circles were made in vitro. Following Hae purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 109 a 150 c 128 g 107 t 1 others
ORIGIN
Query Match 60.8%; Score 370.4; DB 9; Length 495;
Best Local Similarity 94.8%; Pred. No. 7,1e-99;
Matches 383; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Oy 205 GCAGCTGATATACCGAGGGGCTTACCTTGCAGGCGCTGTGCACCGTATACACACCTTC 264
|||||

```

```

Db 1 CGGCGCCGCTTACCGAGGGGCGACCTGCGAGGCGCGCGCTGCACCGTATACACACCTTC 60
Oy 265 AGAGCTGGGCGGCGCCCGCCACCCCAATCCCGGCTATGCGCGAGTATCAAGAGCA 324
Db 61 AGGGCGGCGGCGCCCGCCCGCCGATCCCGGCTACGGCGGAGTATATCAAGAGGCT 120
Oy 325 GTGTATGGCAATTAATTGCTACAGGGTGTGTTACGCTGATACCGCTACGCCACCCACC 384
Db 121 GTGTATGGCAATTAATTGCTACAGGGTGTGTTATGCTGATACCGCTACGCCACCCACC 180
Oy 385 CCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444
Db 181 CCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Oy 445 CACCAACACTTGTCTCAGAGCCCGCCACCTACGCGCTGTGTCATGAATGCTTTGCGCC 504
Db 241 CACCAACACTTGTCTCAGAGCCCGCCACCTACGCGCTGTGTCATGAATGCTTTGCACT 300
Oy 505 TTGACCGATGCCAAGACTAGGAGCCATGCTGATATGTTGGTCTTCTTCTTCTTCTTCT 564
Db 301 TTGACCGATGCCAAGACTAGGAGCCATGCTGATATGTTGGTCTTCTTCTTCTTCTTCT 360
Oy 565 CAGGCTGATATATACCAAGGGGGATACACCGTTTGTCTCATA 608
Db 361 CAGGCTGATATATACCAAGGGGGATACACCGTTTGTCTCATA 404

RESULT 4
LOCUS A1186273 391 bp mRNA linear EST 28-OCT-1998
DEFINITION q20b08.x1 Soares, Placenta 80weeks_2NBHP809W Homo sapiens cDNA
clone IMAGE:1724247 3', mRNA sequence.
A1186273
A1186273.1 GI:3736911
EST.
KEYWORDS
ACCESSION A1186273.1 GI:3736911
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 391)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Straubeberg, Ph.D.
Email: rgs@bbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmerich-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnlnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 369.
Location/Qualifiers
FEATURES
SOURCE
1. .391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1724247"
/clone_lib="Soares, Placenta 80weeks_2NBHP809W"
/dev_stage="Two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGGCGCGCGATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo."
BASE COUNT 73 a 129 c 107 g 82 t
ORIGIN
Query Match 58.7%; Score 357.4; DB 9; Length 391;

```

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.livl.gov>
Series: IRAP Plate: 32 Row: 1 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922072
This clone has the following problem: frame shifted.

FEATURES

FEATURES	source
Location/Qualifiers	
1. .4039	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:4815500"	
/tissue_type="Brain", "hippocampus"	
/clone_id="NHT MCC_95"	
/lab_host="DH10B"	
/note="Vector: pBluescript"	
BASE COUNT	1201 a 858 c 876 g 1104 t
ORIGIN	

DEFINITION	UI-R-E0-bv-c-04-0-0-UI_r1 UI-R-E0 Rattus norvegicus cDNA clone.
ACCESSION	BF549922
VERSION	BF549922.1
KEYWORDS	GI:11659652
SOURCE	EST.
ORGANISM	Norway rat.

REFERENCE

REFERENCE 1 (bases 1 to 533)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)

FEATURES

FEATURES	SOURCE	LOCATION/QUALIFIERS
	1..533	/organism="Rattus norvegicus"
		/strain="Sprague-Dawley"
		/db_xref="taxon:10116"
		/clone="UI-R-E0-by-C-04-0-UI"
		/clone_1ib="UI-R-E0"
		/dev_stage="embryonic"
		/lab_host="DH10B (Life Technologies)"
		/note="vector: pTT10-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI. This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
BASE COUNT	134 a	125 c 164 g 109 t 1 others
ORIGIN		

GenCore version 5.1.4_P5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:40:33 ; Search time 856.8475 Seconds
(without alignments)
11515.871 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactaataaaaaagccgctt.....acaaccgttcgctccatcat 609

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST.*
1: em_gscba:*
2: em_gscbm:*
3: em_gscin:*
4: em_gscm:*
5: em_gscov:*
6: em_gscpl:*
7: em_gscro:*
8: em_hic:*
9: gb_esc1:*
10: gb_esc2:*
11: gb_esc3:*
12: gb_esc4:*
13: gb_esc5:*
14: gb_esc6:*
15: em_escfun:*
16: em_escm:*
17: gb_gsc8:*
18: em_gsc8_hum:*
19: em_gsc8_inv:*
20: em_gsc8_pln:*
21: em_gsc8_vrc:*
22: em_gsc8_fun:*
23: em_gsc8_mam:*
24: em_gsc8_mus:*
25: em_gsc8_other:*
26: em_gsc8_pro:*
27: em_gsc8_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487.2	80.0	4039	11	BC026312 Homo sapi
2	423.4	69.5	533	12	BF549922 UI-R-B0-b
3	370.4	60.8	495	9	AI656926 t48e08.x
4	357.4	58.7	391	9	AI186273 qd20b08.x
5	342	56.2	774	10	AV729057 AV729057
6	338.8	55.6	773	10	AV729198 AV729198

7	327.2	53.7	477	10	AW197589	AW197589 xm44602.x
8	327.2	53.7	480	9	AI401040	AI401040 cb27a12.x
9	314	51.6	353	9	AI189407	AI189407 qd01h05.x
10	314	51.6	449	9	AI374891	AI374891 ca60912.x
11	314	51.6	456	9	AI917298	AI917298 cb96509.x
12	299.8	49.2	607	9	AUI45000	AUI45000 AUI45000
13	289.8	47.6	459	9	AI244212	AI244212 q186b11.x
14	277.6	45.6	439	9	AI968176	AI968176 wu14a06.x
15	277.6	45.6	439	10	AW589795	AW589795 tg22d04.x
16	277.6	45.6	439	10	BE501380	BE501380 7d41n09.x
17	275.8	45.3	521	13	BM183349	BM183349 fv10e03.x
18	275.8	45.3	723	12	BG306387	BG306387 fm58d11.x
19	269.6	44.3	367	9	AI095813	AI095813 qd20g11.x
20	261.8	43.0	420	9	AI655094	AI655094 wds7a10.x
21	259.4	42.6	372	9	AI799929	AI799929 wds7a10.x
22	254.6	41.8	427	12	BF223478	BF223478 7q33e01.x
23	247.2	40.6	392	9	AA975235	AA975235 oq36c08.x
24	247.2	40.6	443	9	AA773715	AA773715 af81c05.x
25	228.6	37.5	680	11	AK005186	AK005186 Mus muscu
26	224	36.8	261	14	F21037	F21037 HSPD05E03.H
27	217.4	35.7	353	14	D60520	D60520 HUM114E06A
28	205.4	33.7	379	9	AI843387	AI843387 UI-M-AQ1-
29	198	32.5	266	9	AI175239	AI175239 EST218774
30	197.4	32.4	430	9	AI850237	AI850237 UI-M-BG1-
31	195.8	32.2	425	10	BB750925	BB750925 BB750925
32	191.2	31.4	218	9	AI189257	AI189257 qc89h03.x
33	187	30.7	327	12	BG186689	BG186689 RST5664.A
34	182.6	30.0	246	9	AI351484	AI351484 q105e06.x
35	180.8	29.7	361	9	AI835011	AI835011 UI-M-AM1-
36	177.8	29.2	655	14	BO188308	BO188308 UI-E-EJ1-
37	177.4	29.1	449	9	AI335996	AI335996 qe42e01.x
38	176.8	29.0	451	9	AI193843	AI193843 qe72h11.x
39	175.2	28.8	347	9	AI041587	AI041587 ox56g05.x
40	175.2	28.8	479	9	AI291784	AI291784 qm75b08.x
41	174.2	28.6	283	12	BG185609	BG185609 RST4560.A
42	172.6	28.3	302	14	BM930547	BM930547 UI-E-EJ1-
43	172.6	28.3	973	11	BC033885	BC033885 Homo sapi
44	167.6	27.5	211	10	AV336519	AV336519 AV336519
45	166.8	27.4	457	9	AI193117	AI193117 qe69h03.x

ALIGNMENTS

RESULT 1
LOCUS BC026312 4039 bp mRNA linear HTC 08-APR-2002
DEFINITION Homo sapiens, clone IMAGE:4015500, mRNA.
ACCESSION BC026312
VERSION BC026312.1 GI:20070932
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiyaki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LUNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.hngc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu

REMARK COMMENT

THIS PAGE BLANK (USPTO)

XX Mus musculus.
OS DE20103510-U1.
XX 07-JUN-2001.
XX 28-FEB-2001; 2001DE-2003510.
XX 02-DEC-1999; 99DE-1058160.
XX (LION-) LION BIOSCIENCE AG.
XX WPI; 2001-368570/39.
XX
XX Gene library containing sequences with specific 3'-ends and no polyA
PT tail, encoding proteins involved in a wide range of cellular processes
PT
PS Claim 15; Page 174; 251pp; German.
XX
XX This invention describes a novel gene library (A) comprises a gene
CC sequence (or its part) encoding a protein involved in amino acid
CC synthesis, cellular/energy metabolism, metabolism of
CC fatty acids/phospholipids, synthesis or breakdown of
CC purines/pyrimidines/nucleosides/nucleotides, DNA
CC replication/transcription/translation, or is a transport/binding protein.
CC (A) are produced that correspond to the 3'-end of mRNA but without the
CC polyA tail. They can be prepared more efficiently and with less effort
CC than conventional libraries. AAK53436-AAK54275 represent fragments of the
CC gene library described in the method of the invention.
XX
SQ Sequence 327 BP; 71 A; 65 C; 84 G; 107 T; 0 other;
Query Match 21.9%; Score 133.6; DB 22; Length 327;
Best Local Similarity 95.8%; Pred. No. 8.2e-31;
Matches 159; Conservative 0; Mismatches 4; Indels 3; Gaps 2;
Qy 447 CCACACACTTGTCTCCA--GCCCCACCTACGCGCTTGTGTCATGATGC-TTTTGGCGC 503
Db 323 CCACACACTTGTCTCCAAGCCGCCCTACGCGCTTGTGTCATGATGCTTTTGGCGC 264
Qy 504 CTTGACCGATGCGCAAGACTAGAGCCATGCTGATGATGCTGCTCTTTCTTCAAT 563
Db 263 CTTGACCGATGCGCAAGACTAGAGCCATGCTGATGATGCTGCTCTTTCTTCAAT 204
Qy 564 GCAGGCTATATATACCAAGGGGATACACCGCTTTGCTCATAT 609
Db 203 GCAGGCTATATATACCAAGGGGATACACCGCTTTGCTCATAT 158
RESULT 15
AAK53229/c
ID AAK53229 standard; cDNA: 1164 BP.
XX
XX AAK53229;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 2758.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dimaac RT, Adundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX P-PSDB; AAM80096.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
PT
PS Claim 1; Page 4968; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, immunomodulatory activity and
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 1666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;
Query Match 21.2%; Score 129; DB 22; Length 1164;
Best Local Similarity 90.2%; Pred. No. 3.6e-29;
Matches 138; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 1 ATGACTAATAAAGGCGGTGAGCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
Db 272 ATGACAAATTAAGAGCCGCTCAACCTTATACAAATGCTGGAATTAATCCAGTTGTG 213
Qy 61 GCGCGGCTCTACAGCCCGGACTTCTATGACAGCAAGGCTGTTGTGCGCAACCAAG 120
Db 212 GGTGAGTCTACAGTCCGCAATTTCTATGACAGCAAGGCTGTTGTGCGCAACCAAG 153
Qy 121 GAGGATTTCCATGATGATGAGTGGCCCGCCAGTTCA 153
Db 152 GAGGATTTCCATGATGATGAGTGGCCCGCCAGTTCA 120

Search completed: March 15, 2003, 12:51:58
Job time : 153.798 secs

CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 939 BP, 148 A, 248 C, 297 G, 246 T, 0 other;

Query Match 22.1%; Score 134.4; DB 22; Length 939;
Best Local Similarity 68.0%; Pred. No. 7.3e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;

QY 166 TCTGCAATGCTGGCTTTCCATATCCGGCCGCACTGTGCAAGTCATACCGAGGGCT 225
DB 417 TATGCAAGTACCGGGTCCCTTACCCACACCGGCAAGCGTTGCTTACCGGGCGCA 358
QY 226 CACCTTCGAGGCGGTGGTGGACCGGTATCAACCTTGAGAGCTGGGGGCCCCACCC 285
DB 357 CATCTTCGGGGGCGGGGCGGTATTAATCAATTTGGGGCTGGCGACCCCAACC 298
QY 286 CCAATCCCGGCTATGCGGAGTAGTATCAAGACCAAGTATGGAATTAATTGCTA 345
DB 297 CCCATCCGCACTTACGAGCGGTGCGTATCAGAGATGATTTATGTCCTGA---GATT 241
QY 346 CAGGTGTTAGCTGATACCGCTAGCCGACCCCAACCCCTGCACTGCTGCTTAC 405
DB 240 TATGGAGGCTACGAGCGCTACAGATAGCTCAGC---CCGTCGACGGGGGAGGCTAC 184
QY 406 AGTGACAGTTAGGAGAGATTTA---TGCTGCCGACCCCTACACACACACTGTCTCA 462
DB 183 AGCGACAGTTAGCGGAGAGCTTACGAGCTGCGGACCCGTCATACACCAATCGGGCCC 124
QY 463 GCCCCACCTACGCGGTGGTGCATGATGCTTTGGCCCTT 506
DB 123 GCGGCGACCTACAGCATTTGGAACCATGTGAACCTTCCACCGTT 80

RESULT 13

AA159248
ID AA159248 standard; cDNA: 1011 BP.

AC AA159248;

XX 22-OCT-2001 (first entry)
DT
XX
DE Human polynucleotide SEQ ID NO 1451.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX
XX

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX MPI: 2001-442253/47.
XX P-PSDB: AAM40092.

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX
PS Claim 1; SEQ ID NO 1451; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 1011 BP; 260 A; 314 C; 267 G; 170 T; 0 other;

Query Match 22.1%; Score 134.4; DB 22; Length 1011;
Best Local Similarity 68.0%; Pred. No. 7.5e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;

QY 166 TCTGCAATGCTGGCTTTCCATATCCGGCCGCACTGTGCAAGTCATACCGAGGGCT 225
DB 615 TATGCAAGTACCGGGTCCCTTACCCACACCGGCAAGCGTTGCTTACCGGGCGCA 674
QY 226 CACCTTCGAGGCGGTGGTGGACCGGTATCAACACTTGAAGCTGGGGGCCCAACC 285
DB 675 CATCTTCGGGGGCGGGGCGGTATTAATCAATTTGGGGCTGGCGACCCCAACC 734
QY 286 CCAATCCCGGCTATGCGGAGTAGTATCAAGACCAAGTATGGAATTAATTGCTA 345
DB 735 CCCATCCGCACTTACGAGCGGTGCTGATCAGATGATTTATGTCCTGA---GATT 791
QY 346 CAGGTGTTAGCTGATACCGCTACCGCCGACCCCTGCCACTGCTGCTTAC 405
DB 792 TATGGAGGCTACGAGCGCTTACAGATAGCTTACG---CCGTCGACGGGGGAGGCTCA 848
QY 406 AGTGACAGTTAGGAGAGATTTA---TGCTGCCGACCCCTTACACACACACTGTCTCA 462
DB 849 AGCGACAGTTAGCGGAGAGTCTACGAGCTGCGGACCCGTCATACACATCGGGCCC 908
QY 463 GCCCCACCTACGCGGTGGTGCATGATGCTTTGGCCCTT 506
DB 909 GCGGCGACCTACAGCATTTGGAACCATGTGAACCTTCCACCGTT 952

RESULT 14

AAK54001/C
ID AAK54001 standard; cDNA: 327 BP.

AC AAK54001;

XX 16-NOV-2001 (first entry)

XX Murine transcription associated protein encoding cDNA SEQ ID 566.

XX Murine; liver; gene library; amino acid synthesis; binding protein;
KM cell metabolism; energy metabolism; fatty acid metabolism; synthesis;
KM phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;
KM replication; transcription; translation; transport protein; ss.

KM Human; nootropic; immunosuppressant; cyrostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM41878.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 5023; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cyrostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;
XX
Query Match 22.1%; Score 134.4; DB 22; Length 939;
Best Local Similarity 66.0%; Pred. No. 7.3e-31;
Matches 234; Conservative 0; Mismatches 10; Indels 9; Gaps 3;
QY 166 TCTGCAATGCTGCTTTCATATCCGCGCACTGCTGAGCTGATACGAGGGGCT 225
DB 417 TATGAGAGGAGGGGTTCCCTACCCACACCGGCAAGCGCTTACCGGGGCGCA 358
QY 226 CACCTTGAGGCGGTGTGCGACCGCTGACAAACCTTCAGAGCTGCGCGCCCGCCACCC 285
DB 357 CATCTTCCGGGCGCGGGCGCGCGCTGTATATATCATTTCCGGGCTGCCACCCCGCCACCC 298
QY 286 CCATTCGGGCGCTATGGGCGAGTAGTATCAAGGCGAGTATGGCAATTAATTGCTA 345
DB 297 CCGATCCGCACTTACGGGCGCGCGCTGTATCAAGATGATTTATGTGTGA--GATT 241

QY 346 CAGGTGTGTTACGCTGCACTACCGCCAGCCACCCCTGCGACTGCTGCTCCCTAC 405
DB 240 TATGAGAGGAGCGGAGCGCTTACAGATATCCCTAGC---CCGCTGACGGCGGAGCCTTAC 184
QY 406 AGTGACAGTTACGAGCAGATTTA---TSGTCCGACCCCTACACACACACTTGCTCA 462
DB 183 AGCGACAGTTACGCGCAGAGTCTACGCACTGCCGACCCGTCATCATACACCATCGGGCC 124
QY 463 GCCCCACCTTACGGCGGTGTGTCATGAATGCTTTGGCCCTT 506
DB 123 GCGGCACTTACAGCATGGAACCATGTGAACCTTCCACCGTT 80
RESULT 12
AA161035/C
ID AA161035 standard; cDNA; 939 BP.
XX
AC AA161035;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5024.
XX
KW Human; nootropic; immunosuppressant; cyrostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM41879.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 5024; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cyrostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIP
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 567 BP; 92 A; 151 C; 169 G; 155 T; 0 other;
Query Match 22.1%; Score 134.4; DB 23; Length 567;
Best Local Similarity 68.0%; Pred. No. 5.9e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;
QY 166 TCTGGAATGCTGCTTCCATATCCGGCCGCTGCGAGCTGCATACCGAGGGCT 225
DB 417 TATGACAGTACGGGGTTCCTTACCCACACCGGACAGCGGTGCTTACCGGGCGCA 358
QY 226 CACCTTCGAGCGCGTGTGCGACCGGTGTAACAACCTTCAGAGCTGCGGCGCCACCC 285
DB 357 CATCTTCGGGGCGCGGGCGCGGTGTATTAATACATTTGGGGCTGCGCCACCCACCC 298
QY 286 CCAATCCCGGCTATGCGGAGTGTATCAAGAGCCAGTGTATGCAATTAATTGCTA 345
DB 297 CCAATCCCGGCTATGCGGAGTGTATCAAGAGCCAGTGTATGCTGCTGA---GATT 241
QY 346 CAGGTTGTTAGCTGCATACCGGTACGCGCCAGCCCGCTGCGACCTGCTGCTTAC 405
DB 240 TATGAGAGGTACGACGCTTACAGATAGCTCAGC---CCGCTGACGCGCGGAGCTAC 184
QY 406 AGTGACAGTTACGACGAGTTA---TGCTGCGACCCCTACACACACACACTTGCTCA 462
DB 183 ACGGACAGTTACGACGAGTGTACGAGCTGCGACCGGTACCATACACACATCGGGCC 124
QY 463 GCCCCACCTACGCGGTGTGCTGATGAATGCTTTGGCGCCTT 506
DB 123 GCGGCGACCTACAGCATGTGAACAATGTAACCTTCCACCGTT 80
RESULT 10
AA159249
ID AA159249 standard; cDNA; 918 BP.
XX
AC AA159249;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1452.
XX
XX Human; nocotropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PR 09-JUL-2000; 2000US-0598042.
XX
PR 19-JUL-2000; 2000US-0620312.
XX
PR 03-AUG-2000; 2000US-0653450.
XX
PR 14-SEP-2000; 2000US-0662191.
XX
PR 19-OCT-2000; 2000US-0693036.
XX
PR 29-NOV-2000; 2000US-0727344.

XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QH, Zhou F, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR P-PSDB; AAM40093.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS
PS Claim 1; SEQ ID NO 1452; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nocotropic,
CC immunosuppressant and cyostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 918 BP; 235 A; 300 C; 235 G; 148 T; 0 other;
Query Match 22.1%; Score 134.4; DB 22; Length 918;
Best Local Similarity 68.0%; Pred. No. 7.2e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;
QY 166 TCTGGAATGCTGCTTCCATATCCGGCCGCTGCGAGCTGCATACCGAGGGCT 225
DB 522 TATGACAGTACGGGGTTCCTTACCCACACCGGACAGCGGTGCTTACCGGGCGCA 581
QY 226 CACCTTCGAGCGCGTGTGCGACCGGTGTAACAACCTTCAGAGCTGCGGCGCCACCC 285
DB 582 CATCTTCGGGGCGCGGGCGCGGTGTATTAATACATTTGGGGCTGCGCCACCCACCC 641
QY 286 CCAATCCCGGCTATGCGGAGTGTATCAAGAGCCAGTGTATGCAATTAATTGCTA 345
DB 642 CCAATCCCGGCTATGCGGAGTGTATCAAGAGCCAGTGTATGCTGA---GATT 698
QY 346 CAGGTTGTTAGCTGCATACCGGTACGCGCCAGCCCGCTGCGACCTGCTGCTTAC 405
DB 699 TATGAGAGGTACGACGCTTACAGATAGCTGAC---CCGCTGACGCGCGGAGCTTAC 755
QY 406 AGTGACAGTTACGACGAGTTA---TGCTGCGACCCCTACACACACACTTGCTCA 462
DB 756 ACGGACAGTTACGACGAGTGTACGAGCTGCGACCGCTACATACACCATCGGGCC 815
QY 463 GCCCCACCTACGCGGTGTGCTGATGAATGCTTTGGCGCCTT 506
DB 816 GCGGCGACCTACAGCATGTGAACAATGTAACCTTCCACCGTT 859
RESULT 11
AA161034/C
ID AA161034 standard; cDNA; 939 BP.
XX
AC AA161034;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5023.
XX

|||||
Db 241 CAGCCCCCCTACGGCGCTGTCATGAACTTTTGCACCTTTGACTGATGCCAAGA 182
Qy 521 CTAGAGCCATGCTGATGATGAGTGTCTGTTCTTTTCATATGACAGCTAGTATACC 580
Db 181 CTAGAGCCATGCTGATGATGAGTGTCTGTTCTTTTCATATGACAGCTAGTATACC 122
Qy 581 AAGGGGATACACCGTTTGTCTCAT 608
Db 121 GAGGGGATACACCGTTTGTCTCAT 94
RESULT 8
AAS93632 standard; cDNA; 406 BP.
XX ID AAS93632 standard; cDNA; 406 BP.
XX AAS93632;
AC AAS93632;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #29436.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS MO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG28445.
DR New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 29436; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 406 BP; 94 A; 97 C; 85 G; 102 T; 28 other;

Query Match 31.2%; Score 190; DB 23; Length 406;
Best Local Similarity 97.5%; Pred. No. 4.8e-48;
Matches 193; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 411 CAGTTACGGAGAGTTTATGTCGCGACCCCTACACACACACTGTGTCAGCCCCAC 470
Db 42 CAGTTACGGAGAGTTTATGTCGCGACCCCTACACACACACTGTGTCAGCCCCAC 101
Qy 471 CTAGAGCGTTGTCATGAAATGCTTTTGCGCCCTTGACCGATCCAGATAGAGCCA 530
Db 102 CTAGAGCGTTGTCATGAAATGCTTTTGCGACCTTGATGATGCCAAGATAGAGCCA 161
Qy 531 TGTGATGATGAGTGTGCTCTGTTCTTTCTTCATTCAGGCTAGTATATACCAAGGGGATA 590
Db 162 TGTGATGATGAGTGTGCTCTGTTCTTTCTTCATTCAGGCTAGTATATACCAAGGGGATA 221
Qy 591 CAACCGTTTGTCTCAT 608
Db 222 CAACCGTTTGTCTCAT 239
RESULT 9
AAS66245/C
XX ID AAS66245 standard; cDNA; 567 BP.
XX AAS66245;
AC AAS66245;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #2049.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS MO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG02058.
DR New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 2049; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in

XX	AA93633;
AC	
XX	
DT	13-FEB-2002 (first entry)
DE	DNA encoding novel human diagnostic protein #29437.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217. 23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dymanac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73. P-FSDB; ABG29446.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
XX	
SB	Claim 1; SEQ ID No 29437; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, or for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantifying a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. A5654197-A594564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

Query Match	49.4%	Score 300.8	DB 82	Length 481
Best Local Similarity	87.2%	Pred. No. 5.9e-23		
Matches 354; Conservative	0	Mismatches 47	Indels 5	Gaps 2
SQ Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;				

QY 205 GGAGCTGCATACCGAGGGGCTTCACCTTGTGAGAGCGTGTGCACCGGTGACACACCTTC 264

Db 1 GGGGCCCCCTACCGAGGGGGCCACTGTGAGAGCGCGGTGCACCGGTGTACMACACCTTC 60

QY 255 AGAGCTGCGGCGCCCCCACCACCCCAATCCCGGCTTATGCGGAGTAGTGTATCAGAGCCA 324

Db 61 AGGGCCGCGGCGCCCCCCCCCGATCCCGGCTTACGGCGGTGTTGTTTACAGAGATGGA 120

QY 325 GGTATGSCAAATTAATTCTACAGGAGTGTACGTGTATACCGTACGCCACGCCACC 384

Db	121	TTTATATG- --TGAGACATTTAATGGTGTTATGCTGCATAACGGCTACGGCCAGCCATACC	1778
Oy	385	CCTGCACCTGCTGTGCTGCCTACAGTACAGTTACGGAGAGTTTATG- -CTGCCGAGCCCT	4424
Db	178	CCTGCACCTGCGCTGCTGCTCTACAGTACAGTTACGGAGAGTTTATGACATTTTCCCCCT	2377
Oy	443	ACCAACAACAACCTTGTCTCCAGCCCCCACTACGGCGTTGGTGCCATGAATGCTTTTGCGC	5020
Db	238	GTCACCACTCACTGCTCCAGCCCCCACTACGGCGTTGGTGCCATGAATGCTTTTGAC	2977
Oy	503	CTTTGACCGATGCCAAGACTATGAGAGCCATGTCTGATATGTGGGCTCTGTTCTTTCTTCAT	5628
Db	298	CTTTGACCTGATGCCAACAATAAGAGCCATCTGATGATGTGGGCTCTGTTCTTTCTTCAT	3575
Oy	563	TGCAGGCTGATATATACCAAGGGGGATACAAACGTTTGTCTCCATA	608
Db	358	TGCAGGCTGATATATACCAAGGGGGATACAAACGTTTGTCTCCATA	403

RESULT 6
AAH09205/c
ID AAH09205 standard; cDNA; 607 BP.

DT	26-JUN-2001	(first entry)
XX		
DE	Human cDNA clone (3'-primer)	SEQ ID NO:6040.

KM	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss
XX	
OS	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss
OS	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss

PN	EP1074617-A2
XX	
PD	07-FEB-2001.
WV	

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300276.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T,

DR WPI; 2001-318749/34

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
XX Claim 3; SEQ ID 6040; 2537pp + CD ROM, English.

CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,

```
Qy 481 GGTGCATGAATGCTTTTGGCCCTTGACCGATGCCAGACTAGAGCCATGCTGATGAT 540
Db 1570 GGTGCATGAATGCTTTTGGCCCTTGACCGATGCCAGACTAGAGCCATGCTGATGAT 1629
Qy 541 GTGGGCTTCGCTTTCTTCTTCATTTGACGGTGTATATATACCAAGGGGATACACCGTTT 600
Db 1630 GTGGGCTTCGCTTTCTTCTTCATTTGACGGTGTATATATACCAAGGGGATACACCGTTT 1689
Qy 601 GCTCCATA 608
Db 1690 GCTCCATA 1697

RESULT 4
AAA07075
ID AAA07075 standard; cDNA, 2372 BP.
XX
AC AAA07075;
XX
DT 03-JUL-2000 (first entry)
XX
DE cDNA encoding human ataxin-2 binding protein (A2BP).
XX
KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
KW neuronal plasticity; cellular degeneration signal transduction pathway;
KW selective RNA transport; spinocerebellar ataxia type-2;
KW hyperproliferative disorder; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 987..1979
FT FT /*tag= a
FT FT /product= "Human A2BP"
FT
XX WO200012710-A1.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20156.
XX
XX 01-SEP-1998; 98US-0145391.
XX
XX (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
XX Pulst SM, Shibata H;
XX
XX WPI; 2000-237873/20.
XX
XX P-PSDB; AAY81462.
XX
XX Nucleic acids encoding an ataxin-2 binding protein useful for
XX inhibiting the expression of active proteins from the SCA2 gene for the
XX treatment of spinocerebellar ataxia type-2.
XX
XX Claim 6; Page 74-77; 82pp; English.
XX
XX This sequence represents cDNA encoding human ataxin-2 binding protein
XX (A2BP). Nucleotide sequences encoding human A2BP were originally
XX isolated in an adult brain cDNA library using the yeast two hybrid
XX method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
XX function that is encoded by the SCA2 gene located on chromosome 12. SCA2
XX has been linked to the autosomal dominant neurodegenerative disorder
XX spinocerebellar ataxia type-2. Individuals afflicted with the disease
XX exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
XX containing a polyglutamine stretch of about 35-39 residues, whereas that
XX of normal individuals contains approximately 22 contiguous glutamine
XX residues. A2BP and ataxin-2 are components of a cellular degeneration
XX signal transduction pathway. The pathogenic expanded form of ataxin-2 has
XX a higher affinity for A2BP relative to normal ataxin-2; the presence of
XX the expanded form is likely to promote degeneration. A2BP and ataxin have
XX also been found to have a role in gene regulation. The binding of A2BP to
XX ataxin-2 plays an important role in controlling gene expression via the
```

```
CC targeting of transport of specific RNAs, selective RNA transport being
CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
CC in embryonic development. Both ataxin-2 and A2BP are able to bind
CC RNA, and are essential components of the RNA localisation network that
CC establishes cellular polarity in embryogenesis. In highly differentiated,
CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
CC function and are required for neuronal plasticity. A2BP nucleic acids may
CC be used for the recombinant production of A2BP proteins or fragments
CC thereof according to standard methodologies. For example, an A2BP protein
CC with an ataxin-2 or RNA binding capability but no signal transduction
CC function can be used as a dominant negative inhibitor of the cellular
CC degeneration signal transduction pathway. A2BP proteins with a signal
CC transduction function can be used to treat hyperproliferative disorders
CC (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX
SQ Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;

Query Match 70.4%; Score 428.6; DB 21; Length 2372;
Best Local Similarity 82.0%; Pred. No. 8.2e-121;
Matches 542; Conservative 0; Mismatches 64; Indels 55; Gaps 2;

Qy 1 ATGACTAATTAATAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATCACTGTG 60
Db 1572 ATGACAAATTAATAAGACCCCTCAACCTTATACAAATGCTGGAATTAATCACTGTG 1631
Qy 61 GGCGGCGTCTACAGCCCGGACTTCTATGACAGGACGAGTCTGTGCGAGGCCAACCAG 120
Db 1632 GGTCGAGTCTACAGTCCGGAATTTATGACAGGACGAGTCTGTGCGAGGCCAACCAG 1691
Qy 121 GAGGATCTTCATATATACAGTGGCCCGGCTGCTGTATATATCTTCGCAATGCTGGC 180
Db 1692 GAGGATCTTCATATATACAGTGGCCCGGCTGCTGTATATATCTTCGCAATGCTGGC 1751
Qy 181 TTTCATATCCGCGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db 1752 TTTCGATATCCAGGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1811
Qy 241 GGTGCACCGGTATACACACCTTTCAGAGCTGCGGCGCGCGCGCGCGCGCGCGCGCTAT 300
Db 1812 GGTGCACCGGTATACACACCTTTCAGAGCTGCGGCGCGCGCGCGCGCGCGCGCTAC 1871
Qy 301 GCGGAGTATGTATATACAGGACGATGTATGCAATTAATTCATACAGGTGTATACGT 360
Db 1872 GCGGAGTATGTATATACAGGACGATGTATGCAATTAATTCATACAGGTGTATACGT 1929
Qy 361 GCATACCGGTATACAGGACGATGTATGCAATTAATTCATACAGGTGTATACGT 410
Db 1930 GCATACCGGTATACAGGACGATGTATGCAATTAATTCATACAGGTGTATACGT 1989
Qy 411 -----CAGTTACGAGCAGTTT 427
Db 1990 TTGCTTCTGTTGACAGATGAATTTCTGTATACACTGTCAGTTACGAGCAGTTT 2049
Qy 428 ATGCTGCGGACCCCTACACCAACACACTTGTCCAGCCCGCCACTACAGCGCTTGTGCA 487
Db 2050 ATGCTGCGGACCCCTACACCAACACACTTGTCCAGCCCGCCACTACAGCGCTTGTGCA 2109
Qy 488 TGAATGCTTTGGCGCTTGAACGATGCCAGACTAGAGCCATGATGATGAGTGC 547
Db 2110 TGAATGCTTTGGCGCTTGAACGATGCCAGACTAGAGCCATGATGATGAGTGC 2169
Qy 548 TCGTTCTTTCTTCAATTTGACGAGGTATATATACCAAGGGGATATACACCGTTTGTCCAT 607
Db 2170 TCGTTCTTTCTTCAATTTGACGAGGTATATATACCAAGGGGATATACACCGTTTGTCCAT 2229
Qy 608 A 608
Db 2230 A 2230

RESULT 5
AAS93633
ID AAS93633 standard; cDNA, 481 BP.
```

Best Local Similarity 90.0%; Pred. No. 1,66-142;
Matches 547; Conservative 0; Mismatches 58; Indels 3; Gaps 1;
QY 1 ATGATAATAAAGAGCGGTGAACCCCTACACCAATGGCTGGAATTAATCAGTGTG 60
Db 806 ATGACAAATTAAGAGACCGCTACACCTTATACAAATGGCTGGAATTAATCAGTGTG 865
QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGGACGCTGCTGTGCGACGCAACGAG 120
Db 866 GGTGAGCTACAGTCCGGAATTTCTATGACAGGACGCTGCTGTGCGACGCAACGAG 925
QY 121 GAGGATCTTCCATATACAGTGGCCCCAGTTTACTTGTATATACTTCTGCAATGCTTGGC 180
Db 926 GAGGATCTTCCATATACAGTGGCCCCAGTTTACTTGTATATACTTCTGCAATGCTTGGC 985
QY 181 TTTCGATATCCGCGCGCCAGCTGCTGACGTGATACCGAGGGGCTCACCTTGAAGCGCT 240
Db 986 TTCCGATATCCAGCCAGCCAGCCGCGGCGCTTACCGAGGGGCTCACCTTGAAGCGCT 1045
QY 241 GGTGCAACCGTGTACACACCTTTCAGAGCTGCGCGCGCCCGCCCAATCCCGGCTTAT 300
Db 1046 GGTGCAACCGTGTACACACCTTTCAGAGCTGCGCGCGCGCCCGCCCGATCCCGGCTTAC 1105
QY 301 GCGGAGTATGCTATCAAGAGCCAGTATGSCAATTAATGCTACAGGCTGTTACCT 360
Db 1106 GCGGAGTATGCTATCAAGAGCCAGTATGSCAATTAATGCTACAGGCTGTTACCT 1162
QY 361 GCATACCGCTACGCGCCAGCCAGCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 1163 GCATACCGCTACGCGCCAGCCAGCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1222
QY 421 CGAGTTTATGCTGCGGACCCCTACACACACCTTGTCTCCAGCCCGCCCACTTACGCGCTT 480
Db 1223 CGAGTTTATGCTGCGGACCCCTACACACACCTTGTCTCCAGCCCGCCCGATCCGCGCTT 1282
QY 481 GGTGCGATGAATGCTTTGGCGCCCTTGAACGATGCCAAGATGAGAGCCATGCTATGAT 540
Db 1283 GGTGCGATGAATGCTTTGGCACTTGTGACTGATGCCAAGATGAGAGCCATGCTATGAT 1342
QY 541 GTGGGTCGTTCTTCTTCAATGAGGCTAGTATATACCAAGGGGGATPACAAACGTTT 600
Db 1343 GTGGGTCGTTCTTCTTCAATGAGGCTAGTATATACCAAGGGGGATPACAAACGTTT 1402
QY 601 GCTCCATA 608
Db 1403 GCTCCATA 1410
RESULT 3
AAK52245
ID AAK52245 standard; cDNA; 1800 BP.
XX
AC AAK52245;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide seq ID NO 790.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN MO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001MO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Weinman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
XX
PT P-PSDB; AAM79112.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 2643-2645; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78333-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulatory
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
XX
Query Match 82.0%; Score 499.2; DB 22; Length 1800;
Best Local Similarity 90.0%; Pred. No. 1,76-142;
Matches 547; Conservative 0; Mismatches 58; Indels 3; Gaps 1;
QY 1 ATGATAATAAAGAGCGGTGAACCCCTACACCAATGGCTGGAATTAATCAGTGTG 60
Db 1093 ATGACAAATTAAGAGACCGCTACACCTTATACAAATGGCTGGAATTAATCAGTGTG 1152
QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGGACGCTGCTGTGCGACGCAACGAG 120
Db 1153 GGTGAGCTACAGTCCGGAATTTCTATGACAGGACGCTGCTGTGCGACGCAACGAG 1212
QY 121 GAGGATCTTCCATATACAGTGGCCCCAGTTTACTTGTATATACTTCTGCAATGCTTGGC 180
Db 1213 GAGGATCTTCCATATACAGTGGCCCCAGTTTACTTGTATATACTTCTGCAATGCTTGGC 1272
QY 181 TTTCGATATCCGCGCGCCAGCTGCTGACGTGATACCGAGGGGCTCACCTTGAAGCGCT 240
Db 1273 TTCCGATATCCAGCGACCGCGCGGCGCTTACCGAGGGGCTCACCTTGAAGCGCT 1332
QY 241 GGTGCAACCGTGTACACACCTTTCAGAGCTGCGCGCGCCCGCCCAATCCCGGCTTAT 300
Db 1333 GGTGCAACCGTGTACACACCTTTCAGAGCTGCGCGCGCCCGCCCGATCCCGGCTTAC 1392
QY 301 GCGGAGTATGCTATCAAGAGCCAGTATGSCAATTAATGCTACAGGCTGTTACCT 360
Db 1393 GCGGAGTATGCTATCAAGAGCCAGTATGSCAATTAATGCTACAGGCTGTTACCT 1449
QY 361 GCATACCGCTACGCGCCAGCCAGCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 1450 GCATACCGCTACGCGCCAGCCAGCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1509
QY 421 CGAGTTTATGCTGCGGACCCCTACACACACCTTGTCTCCAGCCCGCCCACTTACGCGCTT 480
Db 1510 CGAGTTTATGCTGCGGACCCCTACACACACCTTGTCTCCAGCCCGCCCACTTACGCGCTT 1569

PA (SCIO-) SCIOS INC.
 PI Stanton LW, White RT;
 XX WPI: 2002-010779/01.
 XX P-PSDB; AAU70146.
 DR Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX
 XX Claim 1; Fig 1; 189pp; English.
 PS
 CC The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, testenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AA594745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 CC
 XX
 XX Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;
 SQ
 Query Match 100.0%; Score 609; DB 24; Length 1340;
 Best Local Similarity 100.0%; Pred. No. 3,66-176;
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGACTAATAAAGAGCGCTGAACCCCTACACCAATGCTGGAATTAATCCAGTTG 60
 DB 535 ATGACTAATAAAGAGCGCTGAACCCCTACACCAATGCTGGAATTAATCCAGTTG 594
 QY 61 GCGCGGCTTACAGCCCGCACTTATGAGGAGCGGCTGTTGGCAGGCAACG 120
 DB 595 GCGCGGCTTACAGCCCGCACTTATGAGGAGCGGCTGTTGGCAGGCAACG 654
 QY 121 GAGGATCTTCATGATGACAGTGGCCCACTTCTGTATATATCTTCAATGCTG 180
 DB 655 GAGGATCTTCATGATGACAGTGGCCCACTTCTGTATATATCTTCAATGCTG 714
 QY 181 TTTCATATCCGCGCCCACTGCTGAGCTGCATACCGAGGGCTCACTTGAGGCG 240
 DB 715 TTTCATATCCGCGCCCACTGCTGAGCTGCATACCGAGGGCTCACTTGAGGCG 774
 QY 241 GGTCCGACCGTGAACAACCTTCAAGAGTGGCGCCCGCCCAACCCCAATCCCGCT 300
 DB 775 GGTCCGACCGTGAACAACCTTCAAGAGTGGCGCCCGCCCAACCCCAATCCCGCT 834
 QY 301 GCGGAGTATGATCAAGAGCCGATATGCAATTAATGCTTACAGGGTGTACGCT 360
 DB 835 GCGGAGTATGATCAAGAGCCGATATGCAATTAATGCTTACAGGGTGTACGCT 894
 QY 361 GCATACCGCTACCGCCAGCCCACTGCTGCTGCTGCTCAAGTACAGTACGTA 420
 DB 895 GCATACCGCTACCGCCAGCCCACTGCTGCTGCTGCTCAAGTACAGTACGTA 954
 QY 421 CGAGTTATGCTGCGACCCCTTACCAACACACTTGTCCAGCCCGCCCACTTAC 480
 DB 955 CGAGTTATGCTGCGACCCCTTACCAACACACTTGTCCAGCCCGCCCACTTAC 1014
 QY 481 GGTGCGATGAATCTTTGGCCCTTGAACGATGCAAGATGAGCCATGCTGAT 540
 DB 1015 GGTGCGATGAATCTTTGGCCCTTGAACGATGCAAGATGAGCCATGCTGAT 1074
 QY 541 GTGGGTCTGCTTTCTTCTTATGAGGCTAGTATATACCAAGGGGATACCAACG 600

DB 1075 GTGGGTCTGCTTTCTTCTTATGAGGCTAGTATATACCAAGGGGATACCAACG 1134
 QY 601 GCTCCATAT 609
 DB 1135 GCTCCATAT 1143
 RESULT 2
 AAH13824
 ID AAH13824 standard; cDNA; 1513 BP.
 XX
 AC AAH13824;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 DE Human cDNA sequence SEQ ID NO:10786.
 XX
 XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 PD
 XX 07-FEB-2001.
 PF
 XX 28-JUL-2000; 2000EP-0116126.
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Itoai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 PT
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 10786; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;
 Query Match 82.0%; Score 499.2; DB 22; Length 1513;

GenCore version 5.1.4.g5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:24:02 ; Search time 142.798 Seconds
(without alignments)
9604.255 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143
Perfect score: 609
Sequence: 1 atgactaataaaagccgt.....acaacgcttcctccatcat 609

Scoring table: IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: N_GeneSeq_101002.*
2: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT.*
9: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT.*
10: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT.*
11: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT.*
12: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT.*
13: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT.*
14: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT.*
15: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT.*
16: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT.*
17: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT.*
18: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT.*
19: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT.*
20: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT.*
21: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT.*
22: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT.*
23: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT.*
24: /SID52/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	609	100.0	1340	24	AAS94693
2	499.2	82.0	1513	22	AAH13824
3	499.2	82.0	1800	22	AAK52245
4	428.6	70.4	2372	21	AAAG7075
5	300.8	49.4	481	23	AAH93633
6	299.8	49.2	607	22	AAH09205
7	260	42.7	539	24	ABK34530
8	190	31.2	406	23	AAS93632
9	134.4	22.1	567	23	AAS66245

10	134.4	22.1	918	22	AA159249
C 11	134.4	22.1	939	22	AA161034
C 12	134.4	22.1	939	22	AA161035
C 13	134.4	22.1	1011	22	AA159248
C 14	133.6	21.9	327	22	AAK54001
C 15	129	21.2	1164	22	AAK33229
16	112	18.4	3189	24	ABK34602
17	107	17.6	2118	23	AAS93634
18	73.4	12.1	1252	23	AAS90510
19	63.2	10.4	1506	21	AA246827
20	55.2	9.1	60	24	ABN34672
21	47.8	7.8	578	23	AAK90506
22	45	7.4	413	21	AAK04077
23	44.6	7.3	428	24	ABL68917
24	40.4	6.6	316	14	AAK06035
C 25	35.8	5.9	2949	23	AAK67904
C 26	35.6	5.8	577	20	AAK20782
C 27	35.6	5.8	4403765	22	AA199883
28	35.4	5.8	660	20	AAK37406
29	34.6	5.7	2564	23	ABL06127
30	34.6	5.7	5340	23	ABL06126
31	34.6	5.7	12403	23	ABL07646
32	34.2	5.6	15664	17	AAK09312
33	34.2	5.6	15664	18	AAK70492
34	34.2	5.6	15664	18	AAK6130
35	34.2	5.6	15664	18	AAK51224
C 36	34	5.6	2227	23	ABL09457
C 37	34	5.6	3351	20	AA231552
C 38	34	5.6	3974	23	ABL09456
C 39	33.4	5.5	538	19	AAK6456
C 40	33.4	5.5	538	19	AAK44437
C 41	33.4	5.5	538	20	AAK219347
C 42	33.4	5.5	538	20	AAK21935
C 43	33.4	5.5	2207	20	AAK7135
C 44	33.4	5.5	10732	21	AAK10594
45	33	5.4	1828	22	AAK91366

ALIGNMENTS

RESULT 1	
AAS94693	AAS94693 standard; cDNA; 1340 BP.
XX	
AC	AAS94693;
XX	
DT	12-MAR-2002 (first entry)
XX	
DE	Rat secreted factor DNA clone P0184_D11 #1.
XX	
KW	Rat; secreted factor polypeptide; cardiac disease; kidney;
KW	inflammatory disease; congestive heart failure; myocarditis; asthma; ss;
KW	dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
KW	myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
KW	atherosclerosis; cardiac tumor; glomerulonephritis; nephrotic syndrome;
KW	renal infarction; hereditary nephritis; polycystic kidney disease;
KW	chronic renal failure; renal vein thrombosis; medullary sponge kidney;
KW	rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;
KW	graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
KW	Alzheimer's disease; gene therapy.
XX	
OS	Rattus norvegicus.
XX	
PN	W0200174901-A2.
XX	
PD	11-OCT-2001.
XX	
PF	23-MAR-2001; 2001WO-US09555.
XX	
PR	31-MAR-2000; 2000US-193548P.
XX	
PR	14-MAR-2001; 2001US-0809545.
XX	

Human polynucleoti
Human polynucleoti
Human polynucleoti
Human polynucleoti
Murine transcript
Human polynucleoti
Human cDNA for nov
DNA encoding novel
DNA encoding novel
Human RNA binding
Human spliced tran
DNA encoding novel
Human secreted pro
Kidney cancer rela
Human brain Expres
DNA encoding novel
Polynucleotide seq
Mycobacterium tube
Human secreted pro
Drosophila melanog
Drosophila melanog
Drosophila melanog
Mycobacteriophage
Mycobacteriophage
Mycobacteriophage
NheI-G/SpeI fragme
Drosophila melanog
S. rochei strain E
Drosophila melanog
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Human nerve mutat
Gene encoding a su
Peroxidase Swpa2 p

THIS PAGE BLANK (USPTO)

Job time : 1753.58 secs

```
repeat_region complement(99089, .99381)
/rpt_family="Alu"
repeat_region 101648, .101930
/rpt_family="Alu"
repeat_region 102790, .103056
/rpt_family="Alu"
repeat_region 104010, .104277
/rpt_family="Alu"
repeat_region 104010, .104277
/rpt_family="Alu"
repeat_region complement(111121, .111263)
/rpt_family="MER5"
repeat_region complement(112263, .112476)
/rpt_family="MER30"
repeat_region 112956, .113234
/rpt_family="Alu"
repeat_region 114385, .114657
/rpt_family="Alu"
repeat_region complement(117031, .117194)
/rpt_family="MER20"
repeat_region complement(118464, .118770)
/rpt_family="Alu"
repeat_region 119271, .119549
/rpt_family="Alu"
repeat_region complement(122831, .122940)
/rpt_family="MER41"
repeat_region 123471, .123742
/rpt_family="Alu"
repeat_region complement(129860, .129972)
/rpt_family="Alu"
repeat_region 135235, .135310
/rpt_family="MLT1"
repeat_region 135874, .136157
/rpt_family="Alu"
repeat_region 139775, .140048
/rpt_family="Alu"
repeat_region 141734, .141814
/rpt_family="MIR"
repeat_region 143071, .143352
/rpt_family="Alu"
repeat_region 148286, .148773
/rpt_family="Alu"
repeat_region 150431, .150707
/rpt_family="MER1"
repeat_region 155151, .155387
/rpt_family="Alu"
repeat_region complement(156101, .156213)
/rpt_family="MER20"
repeat_region complement(156336, .156920)
/rpt_family="Alu"
repeat_region complement(159633, .159931)
/rpt_family="Alu"
repeat_region complement(160378, .160658)
/rpt_family="Alu"
repeat_region 161617, .162140
/rpt_family="Alu"
repeat_region complement(163200, .163315)
/rpt_family="MIR"
BASE COUNT 48114 a 34256 c 33738 g 48430 t
ORIGIN
```

```
Query Match 18.7% Score 113.6; DB 9; Length 164538;
Best Local Similarity 96.7% Pred. No. 1,1e-21;
Matches 116; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 489 GAATGCTTTTGGCCCTGACGATGCAAGCTAGAGCCATGCTGATGATGCGGCT 548
Db 61350 GAATGCTTTTGCACCTTGACTGATGCCAAGCTAGAGCCATGCTGATGATGCGGCT 61291
Qy 549 CGTCTTCTTCATTCAGGCTAGATATACCAAGGGGATACCAACGTTTGTCTCCATA 608
Db 61290 CGTCTTCTTCATTCAGGCTAGATATACCAAGGGGATACCAACGTTTGTCTCCATA 61231
```

Search completed: March 15, 2003, 14:22:20

RESULT 15
AC005774/c 164538 bp DNA linear PRI 05-OCT-1998
LOCUS Homo sapiens chromosome 16, BAC clone 2603 (LANL), complete
DEFINITION
AC005774
AC005774
VERSION
KEYWORDS
SOURCE HTG. GI:3702264
ORGANISM Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Rieke, D.O.
TITLE 1 (bases 1 to 164538)
JOURNAL Large Scale Sequence Analysis and Annotation with the Sequence
REFERENCE Comparison Analysis (SCAN) System
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 164538)
REFERENCE Unpublished
AUTHORS Rieke, D.O., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.,
and Deaven, L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 164538)
AUTHORS Rieke, D.O., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.,
and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source location/Qualifiers
1..164538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="2603"
1378..1665
/rpt_family="Alu"
repeat_region complement(2144..2320)
repeat_region /rpt_family="MER21"
repeat_region complement(2198..2416)
repeat_region /rpt_family="MER39"
repeat_region complement(2450..2725)
repeat_region /rpt_family="Alu"
repeat_region complement(3157..3393)
repeat_region /rpt_family="Alu"
repeat_region complement(3524..3792)
repeat_region /rpt_family="Alu"
repeat_region 5731..6020
repeat_region /rpt_family="Alu"
repeat_region complement(7557..7727)
repeat_region /rpt_family="MER4"
repeat_region 8282..8928
repeat_region /rpt_family="LTR8"
repeat_region 10055..10310
repeat_region /rpt_family="TBE1"
repeat_region 10472..10618
repeat_region /rpt_family="L1"
repeat_region 10639..10939
repeat_region /rpt_family="Alu"
repeat_region complement(11109..11357)
repeat_region /rpt_family="Alu"
repeat_region 12050..12312
repeat_region /rpt_family="Alu"
repeat_region complement(14642..14755)

repeat_region /rpt_family="L1"
17429..18370
repeat_region /rpt_family="MER21"
complement(19470..19819)
repeat_region /rpt_family="MER21"
complement(20352..20627)
repeat_region /rpt_family="Alu"
23808..24058
repeat_region /rpt_family="Alu"
27654..28217
repeat_region /rpt_family="L1"
complement(32764..32956)
repeat_region /rpt_family="L1"
complement(33392..33503)
repeat_region /rpt_family="Alu"
34586..34870
repeat_region /rpt_family="Alu"
34936..35489
repeat_region /rpt_family="MLT2B2"
complement(36940..37172)
repeat_region /rpt_family="Alu"
37173..37437
repeat_region /rpt_family="TBE1"
41428..41745
repeat_region /rpt_family="TBE1"
complement(42057..43927)
repeat_region /rpt_family="L1"
43924..44555
repeat_region /rpt_family="L1"
45128..45367
repeat_region /rpt_family="Alu"
45467..45704
repeat_region /rpt_family="Alu"
46192..46578
repeat_region /rpt_family="MER25"
complement(46779..47141)
repeat_region /rpt_family="TBE1"
complement(48265..48550)
repeat_region /rpt_family="Alu"
49630..50221
repeat_region /rpt_family="MER41"
50781..51096
repeat_region /rpt_family="Alu"
complement(51219..51439)
repeat_region /rpt_family="Alu"
51462..51708
repeat_region /rpt_family="MER33"
complement(52042..52321)
repeat_region /rpt_family="Alu"
54001..54092
repeat_region /rpt_family="Alu"
complement(55760..56014)
repeat_region /rpt_family="Alu"
56616..56899
repeat_region /rpt_family="Alu"
complement(66463..66747)
repeat_region /rpt_family="Alu"
complement(66908..67180)
repeat_region /rpt_family="Alu"
complement(70389..70704)
repeat_region /rpt_family="MER1"
complement(72166..72421)
repeat_region /rpt_family="MER33"
72501..72769
repeat_region /rpt_family="Alu"
72843..72956
repeat_region /rpt_family="MER5"
complement(76303..76564)
repeat_region /rpt_family="Alu"
97534..97700
repeat_region /rpt_family="MER3"
complement(98469..98526)
repeat_region /rpt_family="Alu"

Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Hollaway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovac, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L. J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsbury, H., Lofado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Melker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okunolu, G., Oregany, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Siscun, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatok, A., Taber, P., Tameisa, A., Tameisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Tellod, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinsom, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE
Journal
Unpublished
2 (bases 1 to 115860)

REFERENCE
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

AUTHORS
Worley, K.C.

JOURNAL
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Jul 18, 2002 this sequence version replaced gi:20514522.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: GXLC
Center clone name: CH230-34C2
Summary Statistics

Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 99693 bases at least Q40
Consensus quality: 101936 bases at least Q30
Consensus quality: 102743 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1347: contig of 1347 bp in length
* 1348 1447: gap of unknown length
* 1448 2834: contig of 1387 bp in length

2835 2934: gap of unknown length
* 2935 4253: contig of 1319 bp in length
* 4254 4353: gap of unknown length
* 4354 5477: contig of 1124 bp in length
* 5478 5577: gap of unknown length
* 5578 7484: contig of 1907 bp in length
* 7485 7584: gap of unknown length
* 7585 9406: contig of 1822 bp in length
* 9407 9506: gap of unknown length
* 9507 11712: contig of 2206 bp in length
* 11713 11812: gap of unknown length
* 11813 13462: contig of 1650 bp in length
* 13463 13562: gap of unknown length
* 13563 17222: contig of 3660 bp in length
* 17223 17322: gap of unknown length
* 17323 20617: contig of 3295 bp in length
* 20618 20717: gap of unknown length
* 20718 24825: contig of 4108 bp in length
* 24826 24925: gap of unknown length
* 24926 28811: contig of 3886 bp in length
* 28812 28911: gap of unknown length
* 28912 33786: contig of 4775 bp in length
* 33787 33787: gap of unknown length
* 33787 40217: contig of 6431 bp in length
* 40218 40317: gap of unknown length
* 40318 47912: contig of 7595 bp in length
* 47913 48012: gap of unknown length
* 48012 55482: contig of 8470 bp in length
* 55483 55483: gap of unknown length
* 55483 65582: gap of unknown length
* 65582 65582: contig of 9622 bp in length
* 65582 66305: gap of unknown length
* 66305 74546: contig of 8242 bp in length
* 74547 74647: gap of unknown length
* 74647 86522: contig of 11876 bp in length
* 86523 86622: gap of unknown length
* 86622 100049: contig of 13427 bp in length
* 100050 100149: gap of unknown length
* 100150 115860: contig of 15711 bp in length.

FEATURES
source 1..115860
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-34C2"

BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others

ORIGIN

Query Match 19.6%; Score 119.4; DB 2; Length 115860;
Best Local Similarity 99.2%; Pred. No. 2.2e-23;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 489 GATGCTTTGGCCCTTGACCGATGCAAGACTAGAGCCATGCTGATGATGGGTCT 548
DB 112555 GATGCTTTGGCCCTTGACCGATGCAAGACTAGAGCCATGCTGATGATGGGTCT 112496

QY 549 CGTCTTTTCTTATGACGCTAGATATATCAAGGGGATACACCGTTTGCTCCATA 608
DB 112495 CGTCTTTTCTTATGACGCTAGATATATCAAGGGGATACACCGTTTGCTCCATA 112436

QY 609 T 609

DB 112435 T 112435

RESULT 14
HSFOX14
LOCUS
DEFINITION Homo sapiens hexarionnucleotide binding protein 1 alpha, beta, and gamma isoforms (HRNPBP1) gene, alternatively spliced, exon 14 and partial cds and complete cds.
ACCESSION AF109120.1 GI:19032383
VERSION
KEYWORDS
SEGMENT 14 of 14

QY	61	GGCGGCGTCTACAGCCCGCATCTTCTATATGAGGACAGGCTGTCTTGTGCGCAAGCCAAACG	120
Db	1632	GGTGAGTCTACAGTCCGAAATTCATATGACGACGGCTCTGTGTGCGCAAGCCAAACG	1691
QY	121	GAGGATCTTCCATATACAGTGGCCCAAGTTACTTGTATATCTTGTGCAATGCTGGC	180
Db	1692	GAGGATCTTCCATATACAGTGGCCCAAGTTACTTGTATATCTTGTGCAATGCTGAGG	1751
QY	181	TTTCCATATCCGGCCGCACTGCTGAGCTGCATACCGAGGGGCTCACCTTGCAGCCGT	240
Db	1752	TTCCGCTATCCAGCAGCCACCCCGCGCGCTACCGAGGGGCGCACCTGCGAGGCGCG	1811
QY	241	GGTGCACCGGTATCAACAACCTTCAGAGCTGGGGGGCCCCCAACCCCAATCCCGGCTAT	300
Db	1812	GGTGCACCGGTATCAACAACCTTCAGAGCGGGGGGGCCCCCGCCCGATCCCGGCTAC	1871
QY	301	GGCGGAGTATGTATCAAGAGCCAGTGTATGCAATTAATTGCTACAGGTTGTTACGT	360
Db	1872	GGCGGATGTTGTTTAC--CAGATGGAATTTATGTTGCGAGACATTTATGTTATGCT	1929
QY	361	GCATACCGCTAGGCCAGCCACCCCTGCGCATGCTGCTGCTTACAGTA-----	410
Db	1930	GCATACCGCTAGGCCAGCCAGCTACCCCTGCGCATGCTGCTTACAGTAAGAAATCAG	1989
QY	411	-----CAGTTACGAGCAGTTT	427
Db	1990	TTCCGCTTCTGTTGACAGCAGATGAATAATTTCTTGTAAACCTCTGCGATTTGACGACGATTT	2049
QY	428	ATGCTGCCGACCCCTTACCAACACACATTTGCTTCAGCCCCCACTAGCGGCTGTGTGCCA	487
Db	2050	ATGCTGCCGACCCCTTACCAACACACATTTGCTTCAGCCCCCACTAGCGGCTGTGTGCCA	2109
QY	488	TGAAGCTTTTGGGGCTTGAACGATATGCCAATAGCAGACATGCTGATGATGATGGGTC	547
Db	2110	TGAAGCTTTTGGACCTTGTACTGATGCGCAACATCAGAGCCATCTGATGATGATGGGTC	2169
QY	548	TGCTTCTTCTTCAATGACGAGCTATATATACAGAGGGGATACACCGTTTGGCTCAT	607
Db	2170	TGCTTCTTCTTCAATGACGAGCTATATATACAGAGGGGATACACCGTTTGGCTCAT	2229
QY	608		
Db	2230		
RESULT 12			
LOCUS	HSN803066	1623 bp	linear
DEFINITION	Homo sapiens mRNA; cDNA DKFp547L059 (from clone DKFp547L059).		
ACCESSION	AL713700		
VERSION	AL713700.1	GI:19584415	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 1623)		
AUTHORS	Bloecher, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAR-2002) MIPS, Am Klopferspitze 18a, D-82152 Martinsried, GERMANY		
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ). Email: s.wiemann@dkfz-heidelberg.de; sequenced by GBR (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.		
	This clone (DKFp547L059) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/ .		

FEATURES	SOURCE	Location/Qualifiers
gene	CDS	1..1623 /organism="Homo sapiens" /db_xref="taxon:9606" /map="16p13.3" /clone="DKFZp547L059" /tissue_type="brain" /clone_id="547 (synonym: hfb11). Vector pSport1; host DH10B; sites NotI + SalI" /dev_stage="fetal" 1..1623 /gene="DKFZp547L059" 836..>1623 /gene="DKFZp547L059" /note="FLJ10165 fls" /codon_start=1 /product="hypothetical protein" /protein_id="CAD28449.1" /db_xref="GI:19584416" /translation="MNCERPOLRGNOEAAAPPTMAQPVASAPAPNPDAIPAEVTA RHPAPAPYTGQTTVPHTLNLVPPAQTHSOSPADTSACTVSGATATDPAIDGPP QTSPENTENKSSQPKRLHVNIPFRFDPDRLQMFQPGILVLEIIENRGSRGFG VTENSNADADNRAREKLHVGVEKRKIEVNNATRVMTNKTLVPTMTKLPVWGA VSPFPAAGVLTLLCOANOEGSSMSAPSLVTSAMPGFYPAAATA"
BASE COUNT	356 a 523 c 487 g 257 t	
ORIGIN		
Query Match	28.3%; Score 172.6; DB 9; Length 1623;	
Best Local Similarity	90.6%; Pred. No. 5.1e-39;	
Matches 184; Conservative	0; Mismatches 19; Indels 0; Gaps 0;	
QY	1	ATGACTATAAAAGGCGGTGAACCCCTCAACCAATGCGTGGAAATTAATCACTGTG 60
Db	1421	ATGACAAATATAAAAGCCGTAAACCTTAATCAATAGCTGGAATGAATCACTGTG 1480
QY	61	GGCGGCTCTACAGCCCGCACTTCTATGACAGCAGCGGTCTGTGTGCGAGCCCAACGAG 120
Db	1481	GGTCACTCTACAGTCCCGAATTTCTATGACAGCAGCGGTCTGTGTGCGAGCCCAACGAG 1540
QY	121	GAGGATCTTCATGTACAGTGGGCCCGATTCACTTGATTAATCTTGCAATGCTGGC 180
Db	1541	GAGGATCTTCATGTACAGTGGGCCCGATTCACTTGATTAATCTTGCAATGCTGGC 1600
QY	181	TTTCCATATCCGGCCGCGACATGC 203
Db	1601	TTCCCGATATCCAGCAGCCACCGC 1623
RESULT 13		
AC120661/c		
LOCUS	AC120661	115860 bp DNA linear HTG 23-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-34C2, *** SEQUENCING IN PROGRESS ***,	
ACCESSION	AC120661	
KEYWORDS	HTG; HTGS_PHASET.	
SOURCE	Rattus norvegicus.	
ORGANISM	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE	1 (bases 1 to 115860)	
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooke,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,	

QY 361 GCATACCGCTACGCGCCAGCCACCCCTGCTGCTGCTTACAGTGA----- 410
 Db 1038 GCATACCGCTACGCGCGCTACCCCTGCGCGCGCTGCTACAGTGAAGAAATCAG 1097
 QY 411 -----CGTTACGAGCAAGTT 427
 Db 1098 TTCGCTTCGTGCGAGCAGATGAATTTCTTGTAAACCTTGCAAGTTACGAGAGTTT 1157
 QY 428 ATGTCGCGACCCCTACACACACACACTGCTTCAGACCCCACTACGCGCTTGTGCGCA 487
 Db 1158 ATGTCGCGACCCCTACACACACACTGCTTCAGACCCCACTACGCGCTTGTGCGCA 1217
 QY 488 TGAATGCTTTGCGCGCTTACCGATGCCAAGACTAGAGCCATGCTGATGATGTTGCGTC 547
 Db 1218 TGAATGCTTTGCGCGCTTACCGATGCCAAGACTAGAGCCATGCTGATGATGTTGCGTC 1277
 QY 548 TCGTCTTTCTTCAATGTCAGGCTAGTATATACCAAGGGGATACAAACGTTTGTCTCCAT 607
 Db 1278 TCGTCTTTCTTCAATGTCAGGCTAGTATATACCAAGGGGATACAAACGTTTGTCTCCAT 1337
 QY 608 A 608
 Db 1338 A 1338

RESULT 10
 AF094849 1547 bp mRNA linear PRI 01-MAR-2002

LOCUS AF094849
 DEFINITION Homo sapiens hexaribonucleotide binding protein 1 isoform alpha
 (HNRBP1) mRNA, complete cds.

ACCESSION AF094849
 VERSION AF094849.1 GI:19032365

KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 1547)
 AUTHORS Chen, W., Chu Z.-L., Blough, R.I., Liu, L., Hoppes, B. and Winkelman, J.C.

TITLE Molecular cloning and chromosomal localization of a human brain, heart and skeletal muscle specific RNA binding protein gene homologous to fox-1 in *Caenorhabditis elegans*

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1547)
 AUTHORS Chen, W., Chu Z.-L., Blough, R.I., Liu, L., Hoppes, B. and Winkelman, J.C.

JOURNAL Direct Submission
 REFERENCE Submitted (24-SEP-1998) Internal Medicine/Hematology-Oncology,
 University of Cincinnati College of Medicine, 231 Bethesda Ave.,
 Cincinnati, OH 45267-0508, USA

TITLE Location/Qualifiers

FEATURES
 source 1..1547
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16p13.3"
 1..1547
 /gene="HNRBP1"
 254..1432
 /gene="HNRBP1"
 /note="similar to *Caenorhabditis elegans* fox-1"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform alpha"

protein_id="AAL83405.1"
 /db_xref="GI:19032366"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLGLIQGQEAADDTMAOP
 VASQAPRPNQIPAEVTAHPHAPAEVGTTPVHTLILYPAQTSBOSPADTSA
 OTVSGAPOTDDAAPTDCOPOTOPSENKSPKRIHYSNI PPRPDDSDLOMFGOF
 GKILDEIITFNRGSKGFGFTFENSADDRAREKILHGTVEGKRIEVNATAPMTN
 KKTVPNTNMGKLNIPVGAIVPEFYAGTVLLCOANQESSN:SAASSLVTSAMGEF
 PYPATAAAVRGAIHLRGRITVYNTFRAAAPPPPIPAYGAVVYDEPVYGNKLGGY

AAVRYAOPPTAAANSDSYGRVYADPYHIALAPATPYGVGAMASYRGYNFARY
 BASE COUNT 380 a 448 c 374 g 345 t
 ORIGIN

Query Match 70.6%; Score 429.8; DB 9; Length 1547;
 Best Local Similarity 92.4%; Pred. No. 9.1e-114;
 Matches 452; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 ATGACATAAATAAAGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTG 60
 Db 899 ATGACAAATTAATAAAGACCGCTCAACCTTATACAAATGCTGGAATTAATTCAGTTGTG 958
 QY 61 GCGCGGTCTACAGCCCGGACCTTATGACAGCAAGTGTGTTGTCAGGCCAAGCAG 120
 Db 959 GGTGAGCTACAGTCCGGAATTTATGACAGCAGGTCTGTGTTGTCAGGCCAAGCAG 1018
 QY 121 GAGGATCTTCATATACAGTGGCCCCAGTTCACTTGTATATCTTTCGAAATGCTGGC 180
 Db 1019 GAGGATCTTCATATACAGTGGCCCCAGTTCACTTGTATATCTTTCGAAATGCTGGC 1078
 QY 181 TTTCATATCCGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 Db 1079 TTCCCGTATCCAGCAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1138
 QY 241 GGTCCACCGGTGTACACACCTTCAGAGCTGGGCGGCCCAACCCCAATCCGCGCTAT 300
 Db 1139 GGTCCACCGGTGTACACACCTTCAGAGCTGGGCGGCCCGGCCCGCCGATCCGCGCTAC 1198
 QY 301 GCGGAGTATGTATATACAGAGCGAGTATGCAATTAATTTCTACAGGTGTATACCT 360
 Db 1199 GCGGAGTATGTATATACAGAGCGAGTATGCAATTAATTTCTACAGGTGTATACCT 1258
 QY 361 GCATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 Db 1258 GCATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1318
 QY 421 CGAGTTTATGCTGCGGACCCCTACACACACACTTGTCTCAGGCCCACTTACGCGGTT 480
 Db 1319 CGAGTTTATGCTGCGGACCCCTACACACACACTTGTCTCAGGCCCACTTACGCGGTT 1378
 QY 481 GGTGCGATG 489
 Db 1379 GGTGCGATG 1387

RESULT 11

LOCUS AR134676 2372 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 1 from patent US 6194171.
 ACCESSION AR134676
 VERSION AR134676.1 GI:14123581

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2372)
 AUTHORS Pulst, S.M. and Shibata, H.

TITLE Nucleic acids encoding ataxin-2 binding proteins
 JOURNAL Patent: US 6194171-A 1 27-FEB-2001;
 FEATURES Location/Qualifiers

BASE COUNT 548 a 726 c 646 g 452 t
 ORIGIN

Query Match 70.4%; Score 428.6; DB 6; Length 2372;
 Best Local Similarity 82.0%; Pred. No. 2.1e-113;
 Matches 542; Conservative 0; Mismatches 64; Indels 55; Gaps 2;

QY 1 ATGACATAAATAAAGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTG 60
 Db 1572 ATGACAAATTAATAAAGCGGTCAACCTTATACAAATGCTGGAATTAATTCAGTTGTG 1631

OY	1	ATGACTAATAAAAGCGGTGAACCCCTTACCACTTAGGTGGAAAATTAAATCAGTTTG	60
Db	899	ATGCACAATAAAGAGACCGTCACACCCTTATACAAATGCGTGAAATTAATCAAGTTTG	958
OY	61	GCGCGGCTCTACAGCCCCCGCATCTTATATGACGACACGCTGCTTTGTGCCAGGCCAACG	120
Db	959	GGTGACGCTACAGTCCCAGAAATTATATGACGACACGCTCTGTGTGGCACGCCAACG	1018
OY	121	GAGGATCTTCCATGTACAGTGGGCCCACTTCACTTGTATATACCTTTCGCATATGCTTGGC	180
Db	1019	GAGGATCTTCCATGTACAGTGGCCCCCACTTCACTTGTATATACCTTTCGCATATGCGAGGC	1078
OY	181	TTTCATATATCCCGCGCCCATCTGTGAGCTGTGATATCCGAGGGGATCACCTTGAGAGCCGT	240
Db	1079	TTCCCTGTATCCACGACACCCCGCGCGGCTTACCGAGGGGCGACCTTGCAGAGCCGC	1138
OY	241	GGTGCGACCGTGTACAAACACTTTCAGAGCTGCGGGGCCCCCAACCCCAATCCCGGCTAT	300
Db	1139	GGTGCGACCGTGTACAAACACTTTCAGAGGCGGGGGCCCCCGCCCGATCCCGGCTAC	1198
OY	301	GCGCGAGTATGTATCAAGAGCCCAAGTGTATGGCAATAAATTCTACAGGGTGTATCGCT	360
Db	1199	GCGCGAGTATGTATCAAGAGCCCTGTGTATGGCAATAAATTCTCAGGGTGTATGCT	1258
OY	361	GCATPCCGCTACGCGCACAGCCGCCCTGTGCATGCGTGCCTACAGTA-----	410
Db	1259	GCATPCCGCTACGCGCACAGCTACCCCTGCATGCGCTCTACAGTACGAATAACG	1318
OY	411	-----CAGTTACGACAGATT	427
Db	1319	TTTCCTCTTGTGACAGCATGAATAATTTCTTATACACTCTGCAAGTTACGACAGATT	1378
OY	428	ATGCTGCCGACCCCTTACCAACAACAACCTTGTCTCAGCCCCCACTTACGCGGTGTGCCA	487
Db	1379	ATGCTGCCGACCCCTTACCAACAACAACCTTGTCTCAGCCCCCACTTACGCGGTGTGCCA	1438
OY	488	TGAATGCTTTTGGCGCCCTTGACCAACATGCCAACAACATGAGAGCATGTAATGATGGGCT	547
Db	1439	TGAATGCTTTTGGCACTTGTACTGTATGCCAACAACATGAGAGCATGTAATGATGGGCT	1498
OY	548	TGCTTCTTTCATGTGAGGCTAGTATATTAACAAGGGGATACAACGTTTGTCTCAT	607
Db	1499	TGCTTCTTTCATGTGAGGCTAGTATATTAACAAGGGGATACAACGTTTGTCTCAT	1558
OY	608 A	608	
Db	1559 A	1559	
RESULT_9			
AB060859		1475 bp mRNA linear PRI 13-JUN-2001	
LOCUS		Macaque faecicularis brain cDNA clone:QcFA-11594, full insert sequence.	
DEFINITION			
VERSION		AB060859.1 GI:13874510	
KEYWORDS		oligo capping; fib (full insert sequence).	
SOURCE		Macaque faecicularis adult male temporal lobe right cDNA to mRNA, clone 11b;macaque brain cDNA library QcFA clone:QcFA-11594.	
ORGANISM		Macaque faecicularis	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;	
TITLE		Cercopithecinae; Macaca.	
JOURNAL		Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,	
REFERENCE		Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.	
AUTHORS		Isolation of full-length cDNA clones from macaque brain cDNA	
TITLE		libraries	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 1475)	
AUTHORS		Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.	
TITLE		Direct Submission	
JOURNAL		Submitted (27-Apr-2001) Kazuyuki Hashimoto, National Institute of	

Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khaethi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
 Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 Lab host: TOP10
 Vector: pME185-F13 (Acc.No. AB009864)
 R. Site1: DraIII (CACTGTGTG)
 R. Site2: DraIII (CACCATGTG)
 Description: 1st strand cDNA was primed with an oligo (dnt) primer
 [ATGTCGACCTTTTGTGTTTTTTTTT]; double-stranded cDNA was synthesized
 using specific 5' and 3' primers and amplified by PCR. The PCR
 product was digested with SfiI and size selection was performed to
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
 into distinct DraIII sites of the pME185-F13. XhoI sites just outside
 the DraIII sites can be used to isolate the cDNA insert. Libraries
 were constructed by oligo-capping method
 (Sugano et al., , Institute of Medical Science, University of
 Tokyo).
 Custom primer used for sequencing
 (5' end primer [CTTTCGCTCTTAAGACGTCG]
 3' end primer [CGACCTCGACGCTCAGCAGCA]).
 Location/Qualifiers
 1. 1475
 /organism="Macaca fascicularis"
 /db_xref="taxon:9541"
 /clone="OtrA-11594"
 /sex="male"
 /tissue_type="temporal lobe right"
 /clone_lib="macaque brain cDNA library OtrA"
 /dev_stage="adult"
 90. 1220
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAB68877.1"
 /db_xref="GI:13874511"
 /translation="MEKSGSRWVQGNQEAAPADPTMAQPYASQNPONGIPAEYTT
 APHPHPAYETGQTYVEPHTLNLVPAPRTSDSESPATNQTYSVITQDDAAPTQSG
 QOPQEPSENTENKQPRHLHSNIPEFRFPDPIROMGCGKILIDVITFNERSGKGF
 GFTFENSAADBARARELHGTVEGKIEVNNTATARMNKTKLVNPYTNMKNLPVVG
 AVSPERFAGTVLCOANQMGSSWYSASLVTASMRGFPYPATAAAYRGANHLRG
 RGRVYVTFPAAPRPRIPAYGCVVQDGGVADYGGVAYRVAQPTPATAAYSRBR
 NQFPAADISCTSVNDFEFLPFTTHLLOPPTALVP"
 BASE COUNT 389 a 414 c 363 g 309 t
 ORIGIN
 Query Match 71.1%; Score 433; DB 9; Length 1475;
 Best Local Similarity 82.5%; Pred. No. 1.1e-114;
 Matches 545; Conservative 0; Mismatches 60; Indels 56; Gaps 2;
 Oy 1 ATGACTAATAAAGAGCCGTGAACCCCTTACCAACCAATGCGTGGAAATTAATTCAGATTGTG 60
 Db 601 ATGACAAATAAAGAGACCGTCAACCTTATACAAATGAGCTGGAATTAATTCAGATTGTG 740
 Oy 61 GCGCGGGCTCTACAGCCCGCATTTCTATGACAGGCAAGCGTCTGTTGTGCGACGCCAACAG 120
 Db 741 GGTGGGGCTCTACAGTCGGGAATTTATGACAGGCAAGCGTCTGTTGTGCGACGCCAACAG 800
 Oy 121 GAGGATTTTCATGATACAGTGGCGCCCGCATTTCACTTGATATTACTTGTGCAATGCGTGGC 180
 Db 801 GAGGATTTTCATGATACAGTGGCGCCCGCATTTCACTTGATATTACTTGTGCAATGCGAGGC 860
 Oy 181 TTTTCATATCCGGCGCCGACCTGTCTGACAGTGCATATCCGAGGGGCTCACCTTTCGAGCCGT 240
 Db 861 TTCCCGTATCCAGAGGACCAACCGCGCGCGCCGCTTACCGAGGGGCGACCTTCGAGGCGCG 920
 Oy 241 GGTGGCACCGGTATCAACACCTTCAAGAGCTGGCGCGCCCAACCCCAATCCCGGCTAT 300
 Db 921 GGCAGCACCGGTATCAACACCTTCAAGGCGCGACGCGCCCGCCCGATCCGCGCTAT 980
 Oy 301 GCGCGAGTAGTATCAAGAGCCAGTGTATGGCAATAATGTCTACAGAGGTGTTACGCT 360
 Db 981 GCGCGGTGTTGTTTACAGAGATGATTTTATGG---TGCAGCATTTATGTGTGTTATGCT 1037

ORIGIN

Query Match 88.4%; Score 538.6; DB 10; Length 1363;
 Best Local Similarity 93.9%; Pred. No. 2.2e-145;
 Matches 572; Conservative 0; Mismatches 34; Indels 3; Gaps 1;

QY 1 ATGACTAATATAAAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCAGTTGTG 60
 DB 595 ATGACAAATATAAGAGCTGTCAACCCCTACACCAATGCTGGAATTAATCAGTTGTG 654
 QY 61 GCGCGGCTTACAGCCCGGCACTTCTATGAGGACGAGTCTGTGTGCGCAACCAACG 120
 DB 655 GCGCGGCTTACAGCCCGGCACTTCTATGAGGACGAGTCTGTGTGCGCAACCAACG 714
 QY 121 GAGGATCTTCCATGATGACAGTGGCCCACTTCACTGTATATATCTTCTGCAATGCTGGC 180
 DB 715 GAGGATCTTCCATGATGACAGTGGCCCACTTCACTGTATATATCTTCTGCAATGCTGGC 774
 QY 181 TTTCCATATCCGCGCCGCACTCTGAGCTGATACCGAGGGGCTCACCTTGAAGCCGT 240
 DB 775 TTCCCATATCCGCGCCGCACTCTGAGCTGATACCGAGGGGCTCACCTTGAAGCCGT 834
 QY 241 GGTGCAACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 300
 DB 835 GGTGCAACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 894
 QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGGGTGTTACGCT 360
 DB 895 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGGGTGTTACGCT 951
 QY 361 GCATACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 420
 DB 951 GGTGCAACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 1011
 QY 421 CGAGTTATGCTGCGGCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 480
 DB 1012 CGAGTTATGCTGCGGCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 1071
 QY 481 GGTGCAACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 540
 DB 1072 GGTGCAACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 1131
 QY 541 GTGGGCTGCTTCTTCTTCAATGAGGCTGATATATACCAAGGGGATACCAACCGTTT 600
 DB 1132 GTGGGCTGCTTCTTCTTCAATGAGGCTGATATATACCAAGGGGATACCAACCGTTT 1191
 QY 601 GCTCCATAT 609
 DB 1192 GCTCCATAT 1200

RESULT 5
 AB041596 2002 bp mRNA linear ROD 30-JUN-2000
 LOCUS Mus musculus brain cDNA, clone MNCB-3035, similar to Homo sapiens
 DEFINITION AB041596
 ACCESSION AB041596
 VERSION AB041596.1 GI:7670455
 KEYWORDS f1s (full insert sequence).
 SOURCE Mus musculus (strain:C57BL) adult female cDNA to mRNA,
 clone lib:Sugano mouse brain mncb clone:MNCB-3035.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Osada, N., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and Hashimoto, K.
 TITLE Isolation of full-length cDNA clones from mouse brain cDNA library made by oligo-capping method
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2002)
 AUTHORS Hashimoto, K., Osada, N., Kusuda, J. and Sugano, S.
 TITLE Direct Submission

JOURNAL

Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khashi@nih.go.jp, URL: <http://www.nih.go.jp/yoken/genebank/>, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 URL: <http://www.nih.go.jp/yoken/genebank/>
 COMMENT
 Lab Name: Sugano mouse brain mncb
 Lab host: TOPI0
 Vector: pME18S-FLJ
 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FLJ. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGTGCTCTAAAGCTGCG]; 3' end primer [CGACCTGCACTCGAGCAC]).

FEATURES

source

A part of this sequence is reported in AU067167.

Location/Qualifiers

1..2002
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="MNCB-3035"
 /sex="female"
 /clone_lib="Sugano mouse brain mncb"
 /dev_stage="adult"
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BA95079.1"
 /db_xref="GI:7670455"
 /translation="MNCERQLRNOEAAAAPDTMAOPYASQFAPRONGIPAEYTA
 HPHPAPRYGTVPTGHTLNDYPTQHSQSATSQTVSGTATQDDAPTDGQPO
 TOPSENTESKSPRLHVSNI PFRFPDLMOMGQKSLDVEII FENERSKQFV
 TFENSADARAREKLHGTVEGRKI EVNATAPRYMTKTYNPTNGKLPVVGAVY
 SPDFYAGTVLCOANOGSSSMYSGPSSLSIVTSABMPFPYPAATAAARGLNRGR
 TVYATFRAAAPPRIPAYCGVYDDGPGVADYGGVAYVRAQGPATPAATVAYSSYGR
 VTADPYHTLPAFTYTGVMNMFAPLTDKTRGHADVGLVLSLOASTYRGRYNR
 FAPY"

CDS

BASE COUNT 518 a 589 c 516 g 379 t

ORIGIN

Query Match 88.4%; Score 538.6; DB 10; Length 2002;
 Best Local Similarity 93.9%; Pred. No. 2.2e-145;
 Matches 572; Conservative 0; Mismatches 34; Indels 3; Gaps 1;

QY 1 ATGACTAATATAAAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCAGTTGTG 60
 DB 1219 ATGACAAATATAAGAGCTGTCAACCCCTACACCAATGCTGGAATTAATCAGTTGTG 1278
 QY 61 GCGCGGCTTACAGCCCGGCACTTCTATGAGGACGAGTCTGTGTGCGCAACCAACG 120
 DB 1279 GCGCGGCTTACAGCCCGGCACTTCTATGAGGACGAGTCTGTGTGCGCAACCAACG 1338
 QY 121 GAGGATCTTCCATGATGACAGTGGCCCACTTCACTGTATATATCTTCTGCAATGCTGGC 180
 DB 1339 GAGGATCTTCCATGATGACAGTGGCCCACTTCACTGTATATATCTTCTGCAATGCTGGC 1398
 QY 181 TTTCCATATCCGCGCCGCACTCTGAGCTGATACCGAGGGGCTCACCTTGAAGCCGT 240
 DB 1399 TTCCCATATCCGCGCCGCACTCTGAGCTGATACCGAGGGGCTCACCTTGAAGCCGT 1458
 QY 241 GGTGCAACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 300
 DB 1459 GGTGCAACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 1518
 QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGGGTGTTACGCT 360
 DB 1519 GCGGAGTAGTGTATCAAGAGCCAGTGTATATGG--TGACACATTTATGTGTATGCT 1575

Db	1408	GCTCCATAT	1416
RESULT 3	AF229057	2000 bp	mRNA linear PRI 01-FEB-2002
LOCUS	AF229057		
DEFINITION	Homo sapiens hexaribonucleotide binding protein 1 isoform gamma (HRNBPI) mRNA, complete cds.		
ACCESSION	AF229057		
VERSION	AF229057.1	GI:18461366	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 2000)		
TITLE	Chen, W. and Winklermann, J.C.		
JOURNAL	Direct Submission		
FEATURES	Submitted (28-JUN-2000) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, The Vontz Center for Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA		
source	Location/Qualifiers		
	1..2000		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16p13.3"		
	1..2000		
gene	/gene="HRNBPI"		
	256..1512		
	/gene="HRNBPI"		
	/note="RNA-binding protein"		
	/codon_start=1		
	/product="hexaribonucleotide binding protein 1 isoform gamma"		
	/protein_id="AAL71904.1"		
	/db_xref="GI:18461367"		
	/translation="MLASQVVLHPYGVMPVPAPIPLIGLQNOEAAAAPDTMAQPPVASFAPPPQNGIPAEVYAPHPAPPEYGTQTVPEHTLNLYPAQTHSESPADTSAQTSGATATOTDDAAPTDGQPOPTOSENTENSOPIRLVNSI PERFPDRLRQMGQFGKILDEVEII FNERGSKGFPVTFENSADAREKLHGTVEGRKI EYNNATATAMTNNGKIVNAGKLVNVAVYSPEYAGTVLCOANOEGSSMSPSLVYTSAMPGKPYPATTAANRGAHLRGRGTIVNTPAAPPPIPAHYGVVYQEPYVQKLLQGVNAAVYPAOPTPATTAAYSDSYGRVVAADPTFHALLAPATYGVGAMNAFAPLDTAKTRSHADDELGVLSSTLQASISYRGVNRFAPY"		
BASE COUNT	502 a 505 c 487 g 506 t		
ORIGIN			
Query Match	89.1%	Score 542.4	DB 9; Length 2000;
Best Local Similarity	93.3%	Pred. No. 1.8e-146;	
Matches 567; Conservative	0; Mismatches 41; Indels 0; Gaps 0;		
1	ATGACTAATTAANAAGGCGGTGAACCCCTTACACCAATGCTGGAATAATCACTGTGG 60		
901	ATGACAAATTAANAAGCGCTCAACCTTATACAAATGCTGGAATTTGAATCACTGTG 960		
61	GGCGGCGTCTACAGCCCGGACCTTCTATGCGAGCAGCGTCTTGTGCCAGGCCAACG 120		
961	GGTGCAGTCTACAGTCCCGAATTCTATGACGACAGCGTCTGTGGCCAGGCCAACG 1020		
121	GAGGATCTTCCATGTACAGTGGGCCCACTTCACTTGTATATACCTTTGCCATTCGCTGC 180		
1021	GAGGGATCTTCCATGTACAGTGGGCCCACTTCACTTGTATATACCTTTGCCATTCGCA 1080		
181	TTTCCATATCGGCGCGGCGCTGCTGAGTGCATACCGAGGGGCTCACCTTGAGAGCGGT 240		
1081	TTTCCGATATCCAGACGCCCGCGCGGCGCTTACCGAGGGGCGCACTTGGAGAGCGGC 1140		
241	GGTGCAGCGGTGTACAAACACTTTCAGAGCTGGCGGCGGCCCAACCCCAATCCCGGCTAT 300		
1141	GGTGCAGCGGTGTACAAACACTTTCAGAGCGCGGCGGCGGCCCAACCCCAATCCCGGCTAT 1200		
301	GCGGAGTAGTAGTATATCAGACCGCAGTGTATGGCAATAATTCCTACAGGAGTGTTACGCT 360		

Db	1201	GCGGAGTAGTGTATCAAGAGCCCTGTGTATGCAATAATTGCTGCAAGGTGTATAGCT	1260
Qy	361	GCATACCGCTACGCGCCAGCCACCCCTGCGCACTGCTGCTGCTTACAGTACAGTTACGGA	420
Db	1261	GCATACCGCTACGCGCCAGCCACCCCTGCGCACTGCTGCTGCTTACAGTACAGTTACGGA	1320
Qy	421	CGAGTTTATGCTGCGCAGCCCTTACACCACTTGTGTCCAGCCCCCACTTACGGGCTT	480
Db	1321	CGAGTTTATGCTGCGCAGCCCTTACACCACTTGTGTCCAGCCCCCACTTACGGGCTT	1380
Qy	481	GATGCCATGAATGCTTTTGGCCGCTTGAACGATGCAAGTACAGGACCATGTATGAT	540
Db	1381	GATGCCATGAATGCTTTTGGCCGCTTGAACGATGCAAGTACAGGACCATGTATGAT	1440
Qy	541	GTGGGCTCTCGTTCTTTCTTTCATTCAGAGCTAGTATATACCAAGGGGGATACCAACCGTTT	600
Db	1441	GTGGGCTCTCGTTCTTTCTTTCATTCAGAGCTAGTATATACCAAGGGGGATACCAACCGTTT	1500
Qy	601	GCTCCATA 608	
Db	1501	GCTCCATA 1508	
RESULT 4			
LOCUS	AF107204	AF107204	1363 bp mRNA linear ROD 23-JUN-2000
DEFINITION	Mus musculus ataxin-2-binding protein (A2bp) mRNA, complete cds.		
ACCESSION	AF107204		
VERSION	AF107204.1	GI:8671587	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1363)		
TITLE	Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.		
JOURNAL	A novel protein, A2BP, with RNA binding motif, binds to C-terminal ataxin-2		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1363)		
TITLE	Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.		
JOURNAL	Direct Submission		
FEATURES	Submitted (16-NOV-1998)		
FEATURES	Division of Neurology and Neurogenetics		
FEATURES	Laboratory, Burns and Allen Research Institute, Cedars-Sinai		
FEATURES	Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA		
FEATURES	location/Qualifiers		
FEATURES	1. .1363		
FEATURES	/organism="Mus musculus"		
FEATURES	/db_xref="taxon:10090"		
FEATURES	1. .1363		
FEATURES	/gene="A2bp"		
FEATURES	67. .1203		
FEATURES	/gene="A2bp"		
FEATURES	/codon_start=1		
FEATURES	/product="ataxin-2-binding protein"		
FEATURES	/protein_id="AA178292.1"		
FEATURES	/db_xref="GI:8671588"		
FEATURES	/translation="MAQPYSAQFAPRPONGIPAEYTAHPHPAPREYTGQTTVDHNTLN		
FEATURES	LYPPTQTHRSOSADTSAQGYSGTATQDDAAPTDDGQPOQPSNTESKSSQPKRLHNS		
FEATURES	IPPRFRDDPDLQMGQRFQFKILDVEITPERSKSGFGFTFENSADADAPRRKLNGT		
FEATURES	VVGRKLEVNNAATARWNTNKTYNPATNGKLPNVGAVVSPVIGATVLCCANDGG		
FEATURES	SSWVGSSSLVITVSAMPGFPPYPAATAAARGLHGRGRVTYNTFRRAAPRPIPAY		
FEATURES	GVGVDDGFGYADIGGYAAYRYAOPTPATPAAYSDSYGTVYAADPVHHTLAPAPTVG		
FEATURES	VGMANAPAPLTDAKTRSHDDVGLVLSLQASIVRGYNNRPFAY"		
FEATURES	358. .375		
FEATURES	/gene="A2bp"		
FEATURES	/note="Region: RNA binding motif RNP-2"		
FEATURES	475. .495		
FEATURES	/gene="A2bp"		
FEATURES	/note="Region: RNA binding motif RNP-1"		
BASE COUNT	362 a 396 c 324 g 281 t		

JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scios Inc. (US)
Location/Qualifiers
FEATURES 1. 1340
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN

Query Match 100.0%; Score 609; DB 6; Length 1340;
Best Local Similarity 100.0%; Pred. No. 7.5e-166;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATCACTGTG 60
DB 535 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATCACTGTG 594

QY 61 GCGCGGTCTACAGCCCGCACTTTATGACAGCAGCGTCTGTGTGCCAGGCAACAG 120
DB 595 GCGCGGTCTACAGCCCGCACTTTATGACAGCAGCGTCTGTGTGCCAGGCAACAG 654

QY 121 GAGGATCTTCATGATGACGTGGCCCACTTGAATTAATCTTGCATAGCTGGC 180
DB 655 GAGGATCTTCATGATGACGTGGCCCACTTGAATTAATCTTGCATAGCTGGC 714

QY 181 TTTCATATCCGCGCCCACTGCTGAGCTGATACCGAGGGCTCACTTCGAGGCGT 240
DB 715 TTTCATATCCGCGCCCACTGCTGAGCTGATACCGAGGGCTCACTTCGAGGCGT 774

QY 241 GGTGCGACCGGTATCAACACCTTTCAGAGCTGGCGGCCCCCAATCCCGGCTAT 300
DB 775 GGTGCGACCGGTATCAACACCTTTCAGAGCTGGCGGCCCCCAATCCCGGCTAT 834

QY 301 GCGGAGTGTATCAAGCAGTATGGAATTAATGTTAGGGTGTATCGCT 360
DB 835 GCGGAGTGTATCAAGCAGTATGGAATTAATGTTAGGGTGTATCGCT 894

QY 361 GATACCGCTACGCGCCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 895 GATACCGCTACGCGCCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 954

QY 421 CGAGTTATGCTGCGACCCCTACACCAACACTTGTCTCCAGCCCCCACTTACG 480
DB 955 CGAGTTATGCTGCGACCCCTACACCAACACTTGTCTCCAGCCCCCACTTACG 1014

QY 481 GGTGCGATGAATGCTTTTGGCCCTTACCGATGCCAAGCTAGAGCATGCTGATG 540
DB 1015 GGTGCGATGAATGCTTTTGGCCCTTACCGATGCCAAGCTAGAGCATGCTGATG 1074

QY 541 GTGGGCTGCTTCTTCTCATTTGAGGCTAGATATACCAAGGGGATACACCGTTT 600
DB 1075 GTGGGCTGCTTCTTCTCATTTGAGGCTAGATATACCAAGGGGATACACCGTTT 1134

QY 601 GCTCCATAT 609
DB 1135 GCTCCATAT 1143

RESULT 2
AF191501 1586 bp mRNA linear ROD 01-MAR-2002
LOCUS AF191501
DEFINITION Mus musculus hexaribonucleotide binding protein 1 (Hrbp1) mRNA,
complete cds.
ACCESSION AF191501
VERSION AF191501.1 GI:19032413
KEYWORDS
SOURCE
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1586)
Chen, W. and Winkelman, J. C.
TITLE Direct Submission

JOURNAL Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
Location/Qualifiers
FEATURES 1. 1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN

Query Match 95.3%; Score 580.2; DB 10; Length 1586;
Best Local Similarity 97.0%; Pred. No. 1.8e-157;
Matches 591; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATCACTGTG 60
DB 808 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATCACTGTG 867

QY 61 GCGCGGTCTACAGCCCGCACTTTATGACAGCAGCGTCTGTGTGCCAGGCAACAG 120
DB 868 GCGCGGTCTACAGCCCGCACTTTATGACAGCAGCGTCTGTGTGCCAGGCAACAG 927

QY 121 GAGGATCTTCATGATGACGTGGCCCACTTGAATTAATCTTGCATAGCTGGC 180
DB 928 GAGGATCTTCATGATGACGTGGCCCACTTGAATTAATCTTGCATAGCTGGC 987

QY 181 TTTCATATCCGCGCCCACTGCTGAGCTGATACCGAGGGCTCACTTCGAGGCGT 240
DB 988 TTTCATATCCGCGCCCACTGCTGAGCTGATACCGAGGGCTCACTTCGAGGCGT 1047

QY 241 GGTGCGACCGGTATCAACACCTTTCAGAGCTGGCGGCCCCCAATCCCGGCTAT 300
DB 1048 GGTGCGACCGGTATCAACACCTTTCAGAGCTGGCGGCCCCCAATCCCGGCTAT 1107

QY 301 GCGGAGTGTATCAAGCAGTATGGAATTAATGTTAGGGTGTATCGCT 360
DB 1108 GCGGAGTGTATCAAGCAGTATGGAATTAATGTTAGGGTGTATCGCT 1167

QY 361 GATACCGCTACGCGCCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 1168 GATACCGCTACGCGCCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1227

QY 421 CGAGTTATGCTGCGACCCCTACACCAACACTTGTCTCCAGCCCCCACTTACG 480
DB 1228 CGAGTTATGCTGCGACCCCTACACCAACACTTGTCTCCAGCCCCCACTTACG 1287

QY 481 GGTGCGATGAATGCTTTTGGCCCTTACCGATGCCAAGCTAGAGCATGCTGATG 540
DB 1288 GGTGCGATGAATGCTTTTGGCCCTTACCGATGCCAAGCTAGAGCATGCTGATG 1347

QY 541 GTGGGCTGCTTCTTCTCATTTGAGGCTAGATATACCAAGGGGATACACCGTTT 600
DB 1348 GTGGGCTGCTTCTTCTCATTTGAGGCTAGATATACCAAGGGGATACACCGTTT 1407

QY 601 GCTCCATAT 609
DB 1407 GCTCCATAT 1413

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:46:57 ; Search time 1628.58 Seconds

(without alignments)
10882.841 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactaataaaagccgt.....acaacgcttcgctccatat 609Scoring table: IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_ey.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pac.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_ats.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vit.*
38: em_ay.*
39: em_hngo_hum.*
40: em_hngo_mus.*
41: em_hngo_other.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	609	100.0	1340	6 AX268800	AX268800 Sequence
2	580.2	95.3	1586	10 AF191501	AF191501 Mus muscu
3	542.4	89.1	2000	9 AF229057	AF229057 Homo sapi
4	538.6	88.4	1363	10 AF107204	AF107204 Mus muscu
5	538.6	88.4	2002	10 AB041596	AB041596 Mus muscu
6	499.2	82.0	1513	9 AK001027	AK001027 Homo sapi
7	499.2	82.0	2279	9 AF107203	AF107203 Homo sapi
8	479.4	78.7	3348	9 AF109106	AF109106 Homo sapi
9	433	71.1	1475	9 AB060859	AB060859 Macaca fa
10	429.8	70.6	1547	9 AF094849	AF094849 Homo sapi
11	428.6	70.4	2372	6 AK134676	AK134676 Sequence
12	172.6	28.3	1623	9 HSM803066	AL1713700 Homo sapi
13	119.4	19.6	115860	2 AC120661	AC120661 Rattus no
14	113.6	18.7	2971	9 HSFOX14	AF1091120 Homo sapi
15	113.6	18.7	164538	9 AC005774	AC005774 Homo sapi
16	111	18.2	251	9 HSFOX09	AF1091115 Homo sapi
17	87.4	14.4	336	9 HSA323668	AF1091115 Homo sapi
18	87.4	14.4	617	9 HSA335811	AF1091115 Homo sapi
19	87.4	14.4	653	9 HSA335616	AF1091115 Homo sapi
20	87.4	14.4	675	9 HSA334013	AF1091115 Homo sapi
21	81.8	13.4	374	9 HSFOX08	AF1091114 Homo sapi
22	80.8	13.3	73929	2 AC103207	AC103207 Rattus no
23	80.4	13.2	1134	10 AF387322	AF387322 Mus muscu
24	80.4	13.2	1695	10 AF229055	AF229055 Mus muscu
25	78.2	12.8	208632	2 AC073624	AC073624 Homo sapi
26	77.4	12.7	234	9 HSFOX13	AF1091119 Homo sapi
27	77.2	12.7	209461	2 AC079359	AC079359 Homo sapi
28	75.6	12.4	1538	9 BC013115	BC013115 Homo sapi
29	73.2	12.0	959	10 AF229056	AF229056 Mus muscu
30	70.4	11.6	159565	2 AL591075	AL591075 Mouse DNA
31	69.8	11.5	115860	2 AC120661	AC120661 Rattus no
32	66.8	11.0	1393	9 AY072786	AY072786 Homo sapi
33	66.8	11.0	1555	9 BC025281	BC025281 Homo sapi
34	66.8	11.0	1925	9 AF228058	AF228058 Homo sapi
35	66.4	10.9	1721	10 BC027263	BC027263 Mus muscu
36	66.2	10.9	151	9 HSFOX11	AF1091117 Homo sapi
37	65.2	10.7	1558	9 AK055213	AK055213 Homo sapi
38	63.2	10.4	1876	9 HSP38A20	AL0092266 H. sapien
39	51.2	8.4	347	9 HSFOX07	AF1091113 Homo sapi
40	51.2	8.4	112976	9 AC027683	AC027683 Homo sapi
41	48.2	7.9	156549	2 AC079072	AC079072 Homo sapi
42	45.6	7.5	125020	9 AF429315	AF429315 Homo sapi
43	45	7.4	2397	9 AK091695	AK091695 Homo sapi
44	44.6	7.3	428	6 AX336745	AX336745 Sequence
45	42	6.9	172090	2 AC018477	AC018477 Homo sapi

ALIGNMENTS

```
RESULT 1
AX268800 LOCUS AX268800 1340 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 Stanton, L.W. and White, R.T.
AUTHORS
TITLE Secreted factors
```


RESULT 15
US-09-960-352-10167/C

; Sequence 10167, Application US/09960352
; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 10167

; LENGTH: 382

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 44-LIB188-014-Q1-E1-C8

US-09-960-352-10167

Query Match 2.0%; Score 27; DB 10; Length 382;

Best Local Similarity 100.0%; Pred. No. 0.17;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1294 AAAAAAAAAAAAAAAAAAATAGC 1320

Db 31 AAAAAAAAAAAAAAAAAAATAGC 5

Search completed: March 16, 2003, 02:34:54
Job time : 127.631 secs

```

; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 714
; LENGTH: 308
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 04-LIB188-006-Q1-E1-A7
; US-09-960-352-714

Query Match
Best Local Similarity 100.0%; Pred. No. 0.18; Length 308;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1294 AAAAAAAAAAAAAAAAAAATTAAGC 1320
Db 33 AAAAAAAAAAAAAAAAAAATTAAGC 7

RESULT 11
US-09-960-352-12366/c
; Sequence 12366, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12366
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 53-LIB2809-032-Q1-E1-F2
; US-09-960-352-12366

Query Match
Best Local Similarity 100.0%; Pred. No. 0.18; Length 315;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1294 AAAAAAAAAAAAAAAAAAATTAAGC 1320
Db 31 AAAAAAAAAAAAAAAAAAATTAAGC 5

RESULT 12
US-09-960-352-7832/c
; Sequence 7832, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7832
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 04-LIB188-006-Q1-E1-A7
; US-09-960-352-714
```

```

; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB188-006-Q1-E1-A6
; US-09-960-352-7832

Query Match
Best Local Similarity 100.0%; Pred. No. 0.17; Length 337;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1294 AAAAAAAAAAAAAAAAAAATTAAGC 1320
Db 63 AAAAAAAAAAAAAAAAAAATTAAGC 37

RESULT 13
US-09-960-352-7833/c
; Sequence 7833, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7833
; LENGTH: 344
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB188-007-Q1-E1-A6
; US-09-960-352-7833

Query Match
Best Local Similarity 100.0%; Pred. No. 0.17; Length 344;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1294 AAAAAAAAAAAAAAAAAAATTAAGC 1320
Db 31 AAAAAAAAAAAAAAAAAAATTAAGC 5

RESULT 14
US-09-960-352-11598/c
; Sequence 11598, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11598
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 50-LIB188-011-Q1-E1-E6
; US-09-960-352-11598

Query Match
Best Local Similarity 100.0%; Pred. No. 0.17; Length 349;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1294 AAAAAAAAAAAAAAAAAAATTAAGC 1320
Db 31 AAAAAAAAAAAAAAAAAAATTAAGC 5
```

OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-42

Query Match 2.2%; Score 30; DB 10; Length 1574;
Best Local Similarity 100.0%; Pred. No. 0.0093;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1286 AACACGAAAAAAAAAAAAAAAAAAAA 1315
Db 1534 AACACGAAAAAAAAAAAAAAAAAAAA 1563

RESULT 6

US-09-960-352-289/C
Sequence 289, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Machiagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 289
LENGTH: 289
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 02-LIB2809-015-Q1-E1-A5
US-09-960-352-289

Query Match 2.1%; Score 28; DB 10; Length 289;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1294 AAAAAAAAAAAAAAAAAATAAGC 1321
Db 36 AAAAAAAAAAAAAAAAAATAAGC 9

RESULT 7

US-09-867-701-9672
Sequence 9672, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9672
LENGTH: 177
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-9672

Query Match 2.0%; Score 27; DB 10; Length 177;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1289 CAACGAAAAAAAAAAAAAAAAAAAA 1315
Db 133 CAACGAAAAAAAAAAAAAAAAAAAA 159

RESULT 8
US-09-983-965-591/C
Sequence 591, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Machiagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT FILING DATE: 2001-10-26
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 591
LENGTH: 214
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 02-LIB188-027-Q1-E1-A5
US-09-983-965-591

Query Match 2.0%; Score 27; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1294 AAAAAAAAAAAAAAAAAATAAGC 1320
Db 44 AAAAAAAAAAAAAAAAAATAAGC 18

RESULT 9

US-09-867-701-9696/C
Sequence 9696, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9696
LENGTH: 263
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-9696

Query Match 2.0%; Score 27; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1289 CAACGAAAAAAAAAAAAAAAAAAAA 1315
Db 40 CAACGAAAAAAAAAAAAAAAAAAAA 14

RESULT 10

US-09-960-352-714/C
Sequence 714, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing

PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 28375
LENGTH: 125
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006075.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.67
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
OTHER INFORMATION: EST HUMAN HIT: Z25303.1, EVALUATE 9.00e-61
OTHER INFORMATION: NT HIT: g11431054, EVALUATE 2.00e-64
OTHER INFORMATION: SWISSPROT HIT: P54537, EVALUATE 1.40e+00
US-09-864-761-28375

Query Match 2.4%; Score 32; DB 10; Length 125;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 GGGCAGCCCGACACACCTTGTGAAAACAC 278
Db 9 GGGCAGCCCGACACACCTTGTGAAAACAC 40

RESULT 4
US-09-864-761-11804
Sequence 11804, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11804
LENGTH: 473
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006075.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.67
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
US-09-864-761-11804

Query Match 2.4%; Score 32; DB 10; Length 473;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 GGGCAGCCCGACACACCTTGTGAAAACAC 278
Db 296 GGGCAGCCCGACACACCTTGTGAAAACAC 327

RESULT 5
US-09-925-302-42
Sequence 42, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 42
LENGTH: 1574
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1029)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1076)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1574)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 16:12:39 ; Search time 119.631 Seconds
(without alignments)
7861.696 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340
Sequence: 1 ggcggccgccccgcagacaat.....ggccgaagggtcgcctaga 1340

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 501302 seqs, 350932545 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	1340	10	US-09-809-545A-1
2	70	5.2	2372	10	US-09-794-591-1
3	32	2.4	125	10	US-09-864-761-28375
4	32	2.4	473	10	US-09-864-761-11804
5	30	2.2	1574	10	US-09-925-302-42
6	28	2.1	289	10	US-09-860-352-289
7	27	2.0	177	10	US-09-867-701-9672
8	27	2.0	214	10	US-09-983-965-591
9	27	2.0	263	10	US-09-867-701-9696
10	27	2.0	308	10	US-09-960-352-714
11	27	2.0	315	10	US-09-960-352-12366
12	27	2.0	337	10	US-09-960-352-7832
13	27	2.0	344	10	US-09-960-352-7833
14	27	2.0	349	10	US-09-960-352-11598
15	27	2.0	382	10	US-09-960-352-10167
16	27	2.0	383	10	US-09-960-352-13506
17	27	2.0	388	10	US-09-960-352-6680
18	27	2.0	394	10	US-09-817-318-16
19	27	2.0	411	10	US-09-960-352-3551

C	20	27	2.0	418	10	US-09-960-352-10650	Sequence 10650, A
C	21	27	2.0	419	10	US-09-960-352-13041	Sequence 13041, A
C	22	27	2.0	424	10	US-09-960-352-12123	Sequence 12123, A
C	23	27	2.0	428	10	US-09-924-035A-516	Sequence 516, App
C	24	26	1.9	222	10	US-09-867-701-8979	Sequence 8979, App
C	25	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	26	26	1.9	230	12	US-10-033-528-96	Sequence 96, App
C	27	26	1.9	222	10	US-09-960-352-8309	Sequence 8309, App
C	28	26	1.9	211	10	US-09-867-701-9449	Sequence 9449, App
C	29	26	1.9	386	10	US-09-960-352-13140	Sequence 13140, A
C	30	26	1.9	390	10	US-09-960-352-1976	Sequence 1976, App
C	31	26	1.9	494	9	US-09-954-531-105	Sequence 105, App
C	32	26	1.9	484	10	US-09-867-701-3275	Sequence 3275, App
C	33	26	1.9	528	10	US-09-951-107-12	Sequence 12, App
C	34	26	1.9	631	10	US-09-879-536-361	Sequence 361, App
C	35	26	1.9	1085	10	US-09-925-300-401	Sequence 401, App
C	36	26	1.9	2422	10	US-09-983-965-4296	Sequence 4296, App
C	37	26	1.9	2481	9	US-10-027-049-3	Sequence 1993, App
C	38	26	1.9	2604	9	US-10-027-049-3	Sequence 1, App
C	39	26	1.9	2738	9	US-10-160-865-17	Sequence 17, App
C	40	26	1.9	3635	10	US-09-529-063-54	Sequence 54, App
C	41	26	1.9	378361	9	US-09-901-136-3	Sequence 3, App
C	42	25	1.9	62	10	US-09-983-965-4296	Sequence 4296, App
C	43	25	1.9	77	10	US-09-764-846-282	Sequence 282, App
C	44	25	1.9	95	10	US-09-960-352-8975	Sequence 8975, App
C	45	25	1.9	111	9	US-09-925-299-635	Sequence 635, App

ALIGNMENTS

RESULT 1
US-09-809-545A-1
; Sequence 1, Application US/09809545A
; Patent No. US2002010804A1
; GENERAL INFORMATION:
; APPLICANT: Stanton, Lawrence W.
; APPLICANT: White, R. Tyler
; TITLE OF INVENTION: SECRETED FACTORS
; FILE REFERENCE: SCIOS.017A
; CURRENT APPLICATION NUMBER: US/09/809,545A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1340
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-809-545A-1

Query Match 100.0%; Score 1340; DB 10; Length 1340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGGCGCCGCTTACACATGCTCAGCTTATGCTTAGCGCACTTGCTCCACCCAGA	60
DB	1	GGGCGCCGCTTACACATGCTCAGCTTATGCTTAGCGCACTTGCTCCACCCAGA	60
QY	61	ATGGCATCTCGCAATATACAGGCGCCCTCATCCCATCCGCGCAGACACCGGCGAG	120
DB	61	ATGGCATCTCGCAATATACAGGCGCCCTCATCCCATCCGCGCAGACACCGGCGAG	120
QY	121	CCCACTGTCCCGCCACATTAATCTTGATCTCTTACACAGACGCACTCGAGCAGAG	180
DB	121	CCCACTGTCCCGCCACATTAATCTTGATCTCTTACACAGACGCACTCGAGCAGAG	180
QY	181	CGCTTATACAGGCGACAGCGCTTCTCCGCGACCGCCACAGACATGATGCCGCCCG	240
DB	181	CGCTTATACAGGCGACAGCGCTTCTCCGCGACCGCCACAGACATGATGCCGCCCG	240
QY	241	ACGACGCGCGACCGCCACAGACAACTTGTGAAAAACAGAAAAAAGATCCCGCCAG	300
DB	241	ACGACGCGCGACCGCCACAGACAACTTGTGAAAAACAGAAAAAAGATCCCGCCAG	300

THIS PAGE BLANK (USPTO)

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 371..1708
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 371..1705
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-327-035-3

Query Match      1.9%; Score 26; DB 4; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
    ||||||||||||||||||||||||
Db 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

```

Search completed: March 16, 2003, 02:32:00
 Job time : 99.6285 secs

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,399
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 371..1708
NAME/KEY: mat_peptide
LOCATION: 371..1705
US-08-838-399-3

Query Match 1.9%; Score 26; DB 2; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
|||
DB 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

RESULT 14
US-09-235-839-3
Sequence 3, Application US/09235839
Patent No. 6207799
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,839
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,118
FILING DATE: April 8, 1996
ATTORNEY/AGENT INFORMATION:

NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 371..1705
US-09-235-839-3

Query Match 1.9%; Score 26; DB 4; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
|||
DB 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

RESULT 15
US-09-327-035-3
Sequence 3, Application US/09327035
Patent No. 6368824
GENERAL INFORMATION:

APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/327,035
FILING DATE: 07-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/838,399
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1234
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

US-09-327-035-1
Sequence 1, Application US/09327035
Patent No. 6368824
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
McCaleb Ph.D., Michael L.
Bloomquist Ph.D., Brian T.
Flores-Riveros Ph.D., Jaime R.
Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/327,035
FILING DATE: 07-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/838,399
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 248..1585
NAME/KEY: mac_peptide
LOCATION: 248..1582
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-327-035-1
Query Match 1.9%; Score 26; DB 4; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 12
US-08-630-118A-3
Sequence 3, Application US/08630118A
Patent No. 5919901
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
McCaleb Ph.D., Michael L.
Bloomquist Ph.D., Brian T.
Flores-Riveros Ph.D., Jaime R.
Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

APPLICANT: Flores-Riveros Ph.D., Jaime R.
Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
FILING DATE: April 8, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1234
TELEFAX: (312)715-1000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 371..1705
US-08-630-118A-3
Query Match 1.9%; Score 26; DB 2; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

RESULT 13
US-08-838-399-3
Sequence 3, Application US/08838399
Patent No. 5965392
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
McCaleb Ph.D., Michael L.
Bloomquist Ph.D., Brian T.
Flores-Riveros Ph.D., Jaime R.
Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 248..1582
US-08-630-118A-1

Query Match 1.9%; Score 26; DB 2; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 9

US-08-838-399-1
Sequence 1, Application US/08838399
Patent No. 5965392
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,399
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 248..1585
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 248..1582
US-08-838-399-1

Query Match 1.9%; Score 26; DB 2; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 10

US-09-235-839-1
Sequence 1, Application US/09235839
Patent No. 6207799
GENERAL INFORMATION:

APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,839
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,118
FILING DATE: April 8, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 248..1582
US-09-235-839-1

Query Match 1.9%; Score 26; DB 4; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 11

;; CURRENT FILING DATE: 1998-10-21
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: PERL Program
;; SEQ ID NO 5
;; LENGTH: 1506
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE: -
;; OTHER INFORMATION: 1250374
US-09-176-657-5

Query Match 2.2%; Score 29; DB 3; Length 1506;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 484 ACCGTGTAGAGGCGCGTAAATCGAGGT 512
Db 717 ACCGTGTAGAGGCGCGTAAATCGAGGT 745

RESULT 3
US-09-306-290-6/c

;; Sequence 6, Application US/09306290
;; Patent No. 6221635
;; GENERAL INFORMATION:
;; APPLICANT: Rovera, Giovanni
;; APPLICANT: Mukhopadhyay, Sunil
;; TITLE OF INVENTION: METHODS FOR SOLID-PHASE AMPLIFICATION OF DNA TEMPLATE
;; FILE REFERENCE: 09924-10
;; CURRENT APPLICATION NUMBER: US/09/306,290
;; CURRENT FILING DATE: 1999-05-06
;; NUMBER OF SEQ ID NOS: 43
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 6
;; LENGTH: 40
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Primer RGP
US-09-306-290-6

Query Match 1.9%; Score 26; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1288 ACAACGAAAAAAAAAAAAAAAAAAAA 1313
Db 26 ACAACGAAAAAAAAAAAAAAAAAAAA 1

RESULT 4
US-09-105-542A-12

;; Sequence 12, Application US/09105542A
;; Patent No. 6323329
;; GENERAL INFORMATION:
;; APPLICANT: Bullerdick, Jörn
;; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF GENES OF THE
;; FILE REFERENCE: BOHMA.001C1P
;; CURRENT APPLICATION NUMBER: US/09/105,542A
;; CURRENT FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: PCT/DE96/02494
;; PRIOR FILING DATE: 1996-12-20
;; NUMBER OF SEQ ID NOS: 23
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12
;; LENGTH: 528
;; TYPE: DNA
;; ORGANISM: Homo Sapien
;; FEATURE:
;; NAME/KEY: misc_feature

;; LOCATION: (1)...(528)
;; OTHER INFORMATION: n = A,T,C or G
US-09-105-542A-12

Query Match 1.9%; Score 26; DB 4; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
Db 472 AACGAAAAAAAAAAAAAAAAAAAA 497

RESULT 5
US-08-341-568-1

;; Sequence 1, Application US/08341568
;; Patent No. 5661021
;; GENERAL INFORMATION:
;; APPLICANT: Buchert, Johanna
;; APPLICANT: Silka-aho, Martti
;; APPLICANT: Vilkaril, Liisa
;; APPLICANT: Penttila, Merja
;; APPLICANT: Saloheimo, Anna
;; APPLICANT: Marjatta, Ranna
;; TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
;; TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching lign
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Birch, Stewart, Kolaesch and Birch
;; STREET: PO Box 747
;; CITY: Falls Church
;; STATE: VA
;; COUNTRY: US
;; ZIP: 22040-0747

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/341,568
;; FILING DATE: 22-NOV-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Murphy Jr, Gerald M
;; REGISTRATION NUMBER: 28,977
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 205-8000
;; TELEFAX: (703) 205-8050
;; TELEX: 248345

;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 536 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ORIGINAL SOURCE:
;; ORGANISM: Trichoderma reesei
;; STRAIN: QM9414
US-08-341-568-1

Query Match 1.9%; Score 26; DB 1; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
Db 460 AACGAAAAAAAAAAAAAAAAAAAA 485

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 15, 2003, 16:12:15 ; Search time 75.6265 Seconds
(without alignments)
5433.755 Million cell updates/sec

Title: US-09-809-545a-1

Perfect score: 1340

Sequence: 1 ggcggccggccctgcacacat.....ggccgaagggttcgtaga 1340

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by change to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70	5.2	2372	4 US-09-145-391-1	Sequence 1, Appl
2	29	2.2	1506	3 US-09-176-657-5	Sequence 5, Appl
3	26	1.9	40	4 US-09-306-290-6	Sequence 6, Appl
4	26	1.9	528	4 US-09-105-542A-12	Sequence 12, Appl
5	26	1.9	536	1 US-08-341-568-1	Sequence 1, Appl
6	26	1.9	536	2 US-08-911-020-1	Sequence 1, Appl
7	26	1.9	631	4 US-09-328-111-361	Sequence 361, App
8	26	1.9	2481	2 US-08-630-118A-1	Sequence 1, Appl
9	26	1.9	2481	2 US-08-838-399-1	Sequence 1, Appl
10	26	1.9	2481	4 US-09-235-839-1	Sequence 1, Appl
11	26	1.9	2481	4 US-09-327-035-1	Sequence 1, Appl
12	26	1.9	2604	2 US-08-630-118A-3	Sequence 3, Appl
13	26	1.9	2604	2 US-08-838-399-3	Sequence 3, Appl
14	26	1.9	2604	4 US-09-235-839-3	Sequence 3, Appl
15	26	1.9	2604	4 US-09-327-035-3	Sequence 3, Appl
16	26	1.9	2610	4 US-09-545-814-3	Sequence 3, Appl
17	26	1.9	2610	4 US-09-545-814-3	Sequence 3, Appl
18	26	1.9	2738	2 US-08-795-868-17	Sequence 17, Appl
19	26	1.9	2738	4 US-09-303-069-17	Sequence 17, Appl
20	26	1.9	2738	4 US-09-134-250-17	Sequence 1, Appl
21	26	1.9	3487	4 US-09-303-069-24	Sequence 24, Appl
22	26	1.9	7291	4 US-08-913-942-3	Sequence 3, Appl
23	25	1.9	291	4 US-09-329-796-1	Sequence 5, Appl
24	25	1.9	572	4 US-09-342-653-5	Sequence 15, Appl
25	25	1.9	599	4 US-09-342-647-19	Sequence 13, App
26	25	1.9	608	4 US-09-385-982-183	Sequence 21, Appl
27	25	1.9	903	5 PCT-US95-06406A-21	

28	25	1.9	1074	4 US-09-357-251-23	Sequence 23, Appl
29	25	1.9	1132	3 US-08-894-731-3	Sequence 3, Appl
30	25	1.9	1375	4 US-09-372-422A-37	Sequence 37, Appl
31	25	1.9	1544	2 US-08-837-593-1	Sequence 1, Appl
32	25	1.9	1584	1 US-08-361-920-24	Sequence 24, Appl
33	25	1.9	1584	1 US-08-479-939-24	Sequence 24, Appl
34	25	1.9	1584	1 US-08-483-432-24	Sequence 24, Appl
35	25	1.9	2071	4 US-09-816-088-1	Sequence 1, Appl
36	25	1.9	2885	2 US-08-791-849A-15	Sequence 15, Appl
37	25	1.9	2901	4 US-08-368-776A-4	Sequence 4, Appl
38	25	1.9	2901	5 PCT-US96-00419-4	Sequence 4, Appl
39	25	1.9	3180	1 US-08-480-662-1	Sequence 1, Appl
40	25	1.9	3180	3 US-08-818-190-1	Sequence 1, Appl
41	25	1.9	3180	4 US-09-234-232-1	Sequence 1, Appl
42	25	1.9	3180	5 PCT-US96-09927-1	Sequence 1, Appl
43	25	1.9	3238	4 US-08-123-934A-5	Sequence 5, Appl
44	25	1.9	3238	5 PCT-US94-10080-5	Sequence 5, Appl
45	25	1.9	3410	4 US-09-020-956-110	Sequence 110, Appl

ALIGNMENTS

RESULT 1
US-09-145-391-1
Sequence 1, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulist, Stefan M.
TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145,391
CURRENT FILING DATE: 1998-09-01
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (987) ..(1979)
US-09-145-391-1

Query Match 5.2%; Score 70; DB 4; Length 2372;
Best Local Similarity 100.0%; Pred. No. 6.4e-21;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATCCAAAGACTAGGACCATGCTGATGATGAGTGGTCTGTTCTTTCTTCAATTCAGAGCT 1104
DB 2133 GATCCAAAGACTAGGACCATGCTGATGATGAGTGGTCTGTTCTTTCTTCAATTCAGAGCT 2192

QY 1105 AGTATATACC 1114
DB 2193 AGTATATACC 2202

RESULT 2
US-09-176-657-5
Sequence 5, Application US/09176657
Patent No. 6020164
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Tang, Y. Tom
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Lu, Aina
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
FILE REFERENCE: PF-0611 US
CURRENT APPLICATION NUMBER: US/09/176,657

THIS PAGE BLANK (USPTO)

```

/c1one="UTSM_SM385"
/c1one_lib="UTSM Adult Mouse Skeletal Muscle Library"
/sex="Pooled"
/tissue_type="Diaphragm/Hind limb muscles"
/cell_type="Skeletal muscle"
/dev_stage="2 months"
/lab_host="DH5a"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Mannick); Cloning Technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TAGCTCCACTGATTTCTGACTG--->. Other
information regarding entire library may be found at
http://psa.bwmed.edu/Data/Libraries/microarray_cdna_librar
ies.htm."

BASE COUNT      108 a      104 c      91 g      101 t      4 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 80; DB 12; Length 408;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 ACCCTACACCAATGCTGGAATTAATCCAGTTGGGGGGGCTACAGCCCCGAC 615
    |||
Db 232 ACCCTACACCAATGCTGGAATTAATCCAGTTGGGGGGGCTACAGCCCCGAC 291
    |||

QY 616 TTCTATGACGACGCTGCT 635
    |||
Db 292 TTCTATGACGACGCTGCT 311
    |||

RESULT 15
228900/c
LOCUS          228 bp      mRNA      linear      EST 14-DEC-1993
DEFINITION    HSB87F031 STRATAGENE Human skeletal muscle cDNA library, cat.
ACCESSION     228900
VERSION       228900.1 GI:434570
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 228)
AUTHORS      Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
               M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
               Sebaastiani-Kabakcthis,C. and Tessier,A.
               IMAGE: molecular integration of the analysis of the human genome
               and its expression
               C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
JOURNAL       95277534
MEDLINE
COMMENT       Contact: Genethon
               Genexpress-Genethon
               Genethon Centre de recherche sur le Genome Humain
               1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
               Tel: 33169472800
               Fax: 3316078698
               Email: genexpress@genethon.fr
               etrand(-), single read.
               Location/Qualifiers
                   1..228
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="B7F03"
                       /clone_lib="STRATAGENE Human skeletal muscle cDNA library,
                       cat. #936215."
                       /sex="Female"
                       /tissue_type="skeletal muscle"
                       /dev_stage="19 years"
                       /note="Organ: leg muscle; Vector: Lambda ZAPII; Tissue
                       from female, 19 years old, normal leg muscle. Cloning
                       vector is Lambda ZAPII, in vivo excision from lambda ZAPII

```

```

BASE COUNT      64 a      51 c      48 g      64 t      1 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 70; DB 14; Length 228;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGAGCCATGCTGATGATGGGCTCTTTCTTCATTGACAGCT 1104
    |||
Db 201 GATGCCAAGACTAGAGCCATGCTGATGATGGGCTCTTTCTTCATTGACAGCT 142
    |||

QY 1105 AGTATATACC 1114
    |||
Db 141 AGTATATACC 132
    |||

Search completed: March 16, 2003, 02:29:40
Job time : 1895.21 secs

```



```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-acb-d-03-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(Bonaldi, Lennon and Soares, Genome Research 6: 791-806,
1996)
TAG_LIB=UI-R-Y0
TAG_TISSUE=EYE
TAG_SEQ=CATTG"
BASE COUNT      103 a      81 g      64 t
ORIGIN

```

```

Query Match      10.4%; Score 140; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 200 CCGTCCGGGACCGCCACAGACAGATGATGCGCCCGCCGACGAGCCGACCCCA 259
      |||||||
DB 205 CCGTCCGGGACCGCCACAGACAGATGATGCGCCCGCCGACGAGCCGACCCCA 264
      |||||||
QY 260 CACAACCTTCTGAAAAACAGAAAAACAAGTCCAGCCCAAGCGGTGATGTGTCAACA 319
      |||||||
DB 265 CACAACCTTCTGAAAAACAGAAAAACAAGTCCAGCCCAAGCGGTGATGTGTCAACA 324
      |||||||
QY 320 TCCCTTCGCGTCCGGAT 339
      |||||||
DB 325 TCCCTTCGCGTCCGGAT 344
      |||||||

```

```

RESULT 13
BI988957      600 bp mRNA linear EST 20-DEC-2001
LOCUS      4023-45 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
DEFINITION      mRNA sequence.
ACCESSION      BI988957
VERSION      BI988957.1 GI:17959947
KEYWORDS
SOURCE
ORGANISM      Mus musculus.
                house mouse.
REFERENCE      1 (bases 1 to 600)
AUTHORS      Mu,X., Zhao,S., Pershad,R., Heieh,T.-F., Scarpa,A., Wang,S.W.,
                White,R.A., Beremand,P.D., Thomas,T.B., Gan,L. and Klein,W.H.
TITLE      Gene expression in the developing mouse retina by EST sequencing
                and microarray analysis
JOURNAL      Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE      21671825
COMMENT      Contact: Klein WH
                Department of Biochemistry and Molecular Biology
                University of Texas M.D. Anderson Cancer Center

```

```

Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
Location/Qualifiers
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/cisue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Mammatis); Cloning technique: CUA Cloning (Clontekmp,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGTCCACTGAAATCTGAGTG--->. Other
information regarding entire library may be found at
http://pga.sommed.edu/data/libraries/microarray_cdna_librar
ies.htm."

```

```

BASE COUNT      171 a      124 c      122 g      183 t
ORIGIN

```

```

Query Match      7.2%; Score 97; DB 13; Length 600;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 147; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1023 GAATGCTTTTGCGCCCTTGACCGATGCAAGACTAGAGCCATGCTGATGTTGGTCT 1082
      |||||||
DB 450 GAATGCTTTTGCGCCCTTGACCGATGCAAGACTAGAGCCATGCTGATGTTGGTCT 509
      |||||||
QY 1083 CGTCTCTTCTTCATGACGGTATATATACCAAGGGGATCAACCGTTTGTCTCAT 1142
      |||||||
DB 510 CGTCTCTTCTTCATGACGGTATATATACCAAGGGGATCAACCGTTTGTCTCAT 569
      |||||||
QY 1143 TTAATGATATAAACCATTAACAAACA 1170
      |||||||
DB 570 TTAATGATATAAACCATTAACAAACA 597
      |||||||

```

```

RESULT 14
BG795922      408 bp mRNA linear EST 16-MAY-2001
LOCUS      UTSW_SMB85 UTSW Adult Mouse Skeletal Muscle Library Mus musculus
DEFINITION      UTSW_SMB85 UTSW Adult Mouse Skeletal Muscle Library Mus musculus
ACCESSION      BG795922
VERSION      BG795922.1 GI:14131492
KEYWORDS      cDNA clone UTSW_SMB85, mRNA sequence.
SOURCE
ORGANISM      house mouse.
                Mus musculus.
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 408)
AUTHORS      Gallardo,T.D., Schageman,J.V., Persemlidis,A., Garner,H.R.,
                Williams,R.E. and Shohet,R.V.
TITLE      UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA
                Library
                Unpublished (2001)
JOURNAL      Contact: Schageman JJ
                Shohet/Garner Labs
                University of Texas Southwestern Medical Center
                6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
                Tel: 214 648 1674

```

```

Email: Jeff.Schageman@UTSouthwestern.edu
cDNA library constructed by UTSW as a component of the program for
Genomic Applications (PGA) and the Reynolds Heart Disease
Prevention grants for use in cDNA microarray experiments. Sequence
Quality: Sequence ends were trimmed based on percentage of ambigu
us base calls or 'N's in windowed segments. Sequencing: First-pass
sequencing; ABI Prism 377 sequencer and analysis software.
Seq primer: M13/pUC Reverse.
Location/Qualifiers
1..408
/organism="Mus musculus"
/db_xref="taxon:10090"

```

```

FEATURES
SOURCE

```

ORIGIN

Query Match 11.7%; Score 157; DB 9; Length 361;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 207; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 963 TGCTGCCGACCCCTTACACACACACTTGTCCAGCCCTTACAGCGCTGTGTCAT 1022
 DB 358 TGCTGCCGACCCCTTACACACACACTTGTCCAGCCCTTACAGCGCTGTGTCAT 299
 QY 1023 GAATGCTTTTGGCCCTTGACCGATGCGCAAGATGAGAGCCATGCTGTGTCAT 1082
 DB 298 GAATGCTTTTGGCCCTTGACCGATGCGCAAGATGAGAGCCATGCTGTGTCAT 239
 QY 1083 CGTTCTTTCTTCATGACGCTAGTATATACCAAGGGGATACCAACCGTTTGTCCATA 1142
 DB 238 CGTTCTTTCTTCATGACGCTAGTATATACCAAGGGGATACCAACCGTTTGTCCATA 179
 QY 1143 TTAAATGATAAACCATTAACCAACA 1170
 DB 178 TTAAATGATAAACCATTAACCAACA 151

RESULT 11
 LOCUS BB750925 425 bp mRNA linear EST 16-OCT-2001
 DEFINITION BB750925 RIKEN full-length enriched, pooled tissues, cerebellum,
 etc. Mus musculus cDNA clone G130203B11 3', mRNA sequence.
 ACCESSION BB750925
 KEYWORDS BB750925.1 GI:16155161
 SOURCE EST.
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 425)

REFERENCE
 AUTHORS Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,
 Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
 Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
 Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,
 A., Takahashi,F., Takaku-Akenita,S., Tanaka,T., Tomaru,A., Toya,T.,
 Watanabe,A., Yasunishi,A., Yamamoto,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

JOURNAL COMMENT Unpublished (2001)
 Contact: Yoshinori Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: http://genome.resc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
 M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

FEATURES
 source
 1..425
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G130203B11"
 /clone_1b="RIKEN full-length enriched, pooled tissues,
 cerebellum, etc."
 /note="pooled tissues; (tissue_type=cerebellum,
 dev_stage=16 days neonate, sex=mixed),
 (tissue_type=cerebellum, dev_stage=0 day neonate,
 sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
 sex=male), (tissue_type=whole body, dev_stage=9 days
 embryo, sex=mixed) (tissue_type=liver, dev_stage=13 days
 embryo, sex=mixed)"

BASE COUNT 128 a 109 c 87 g 101 t

ORIGIN
 Query Match 10.9%; Score 146; DB 10; Length 425;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 CAGTTACGACGAGATTATGCTCCGACCCCTTACACACACACTTGTCCAGCCCTTAC 1004
 DB 73 CAGTTACGACGAGATTATGCTCCGACCCCTTACACACACACTTGTCCAGCCCTTAC 132
 QY 1005 CTACGGCGTTGGTGCATGATGATCTTTGGCCCTTGACCGATGCGCAAGCTAGAGCCA 1064
 DB 133 CTACGGCGTTGGTGCATGATGATCTTTGGCCCTTGACCGATGCGCAAGCTAGAGCCA 192
 QY 1065 TGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1090
 DB 193 TGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 218

RESULT 12
 LOCUS A1717457 395 bp mRNA linear EST 10-JUN-1999
 A1717457
 DEFINITION UI-R-Y0-acd-d-03-0-UI-81 UI-R-Y0 Rattus norvegicus cDNA clone
 UI-R-Y0-acd-d-03-0-UI 3', mRNA sequence.
 ACCESSION A1717457
 VERSION A1717457.1 GI:5034713
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 395)

REFERENCE
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.iowa.edu
 Oligo-dt track not found. Not a site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: Clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=No.

FEATURES
 source
 1..395
 /location/Qualifiers
 /organism="Rattus norvegicus"

Qy	1143	TTAAATGATAAACCAATTAACCAACACAGCAAAAACAAACCAAAAACCAAAATACCAAC	1202
Db	164	TTAAATATATAAACCAATTTAAACCAACCAAAACAAACCAAAAACAAAAAACCAAAACCAAC	105
Qy	1203	CTTCCAAATGTGGGAGAGAGAGAAAGCTTCCGAGGCCGAGTGTGGCAACATGCAATG	1262
Db	104	CTTCCAAATGTGGGAGAGAGAGAAAGCTTCCGAGGCCGAGTGTGGCAACATGCAATG	45
Qy	1263	GACATCACTTTAGCAACTCAAGAAACAAAC	1292
Db	44	GACATCACTTTAGCAACTCAAGAAACAAAC	15

RESULT 9					
BFS49922/c					
LOCUS	BFS49922	533 bp	mRNA	linear	EST 12-DEC-2000
DEFINITION	UI-R-E0-dv-C-04-0-UI.#1 UI-R-E0 Rattus norvegicus cdna clone				
ACCESSION	UI-R-E0-dv-C-04-0-UI.5,				
VERSION	BFS49922				
KEYWORDS	BFS49922.1 GI:11659652				
SOURCE	EST.				
ORGANISM	Norway rat.				
	Rattus norvegicus				

REFERENCE 1 (bases 1 to 533)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel.: 319 335 8250
Fax: 319 335 9565
Email: mscores@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
Clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNLN (info@image.llnl.gov). IMAGE ID= 1777166
Seq primer: M3 forward.

```

FEATURES
source
location/Qualifiers
1..533
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-C-04-0-UI"
/clone_1b="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pTR73D-pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT
134 a 125 c 164 g 109 t 1 others
ORIGIN

```

Query Match	11.9%	Score 160;	DB 12;	Length 533;
Best Local Similarity	99.5%	Pred. No. 0;		
Matches 210; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

Qy	904	TACGCCACGCCACCCCTCTGCTCTGCTCTACAGTACAGTTACGAGAGATTAT	963
Db	268	TACGCCACGCCACCCCTCTGCTCTGCTCTACAGTACAGTTACGAGAGATTAT	209
Qy	964	GCTGCCAGCCCTACACACACACTTGTCTCAACCCCCACCTACGGCGTTGTGTGCATG	1023

D _b	208	GCTCCGACCCCTTACCAACAACACTTGCTCAGCCCCCATCATCGGGCGTTGGGCATG	149
O _y	1024	AATGCTTTTGGGCCCTTGACCGAATGCCAAAGATAAGAGCATGTGATGATGGGTCTC	108
D _b	148	AATGCTTTTGGGCCCTTGACCGAATGCCAAAGATAAGATGATGGGTCTC	89
O _y	1084	GTTCTTCCTTCATTGACAGGCTAGATATACC	1114
D _b	88	GTTCTTCCTTCATTGACAGGCTAGATATACC	58

RESULT	10
A1835011/c	
LOCUS	361 bp mRNA linear EST 14-JUL-1999
DEFINITION	UI-M-AM1-afw-h-01-0 UI s1 NIH BMP_MAM_N Mus musculus cDNA clone.
ACCESSION	A1835011
VERSION	A1835011
KEYWORDS	A1835011.1 GI:5469224
SOURCE	EST.
ORGANISM	house mouse. Mus musculus Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	1 (bases 1 to 361)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Chin, H

National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel.: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized amygdala library cDNA library Preparation: M.B. Soares lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 122-168,
>(CAAA)n#simple repeat
Seq primer: M13_Foward
POLYA=yes.

FEATURES	Location/Qualifiers
source	1. .361

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U1-M-AM1-afw-h-01-0-UI"
/clone_lib="NIH_BMAP_MAM_N"
/seq_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH BMAP MAM N library is a normalized library constructed
from mouse amygdala. The tag is a string of 5 nucleotides
present between the Not I site and the Oligo-dT track.
The library was constructed as described by Bonaldo,
Lemon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Ms. Annie Novakovich, Zivic-Miller
laboratories.
TAG_LIB=NIH_BMAP_MAM_N
TAG_TISSUE=amygdala
TAG_SEQ=GTGAG

```

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 430)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: MEST@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a donatide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized spinal cord library cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence:
 116-162, >(CAA)n#Simple_repeat
 Seg primer: M13 Forward
 POLYA=Yes

FEATURES
 source Location/Qualifiers
 1..430
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BG1-ale-e-06-0-UI"
 /clone_1lb="NIH_BMAP_MSC_N"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTZ19D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_MSC_N library is a normalized library constructed from mouse spinal cord. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares. Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
 TAG_L1B=NIH_BMAP_MSC_N
 TAG_TISSUE=spinal-cord
 TAG_SEQ=TCGAA"

BASE COUNT 97 a 85 c 104 g 144 t
 ORIGIN
 Query Match 13.1%; Score 175; DB 9; Length 430;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 945 CAGTACGGACGAGTTATGCTGCGGACCCCTACACACACACCTGCTCAGCCCCAC 1004
 DB 370 CAGTTACGGACGAGTTATGCTGCGGACCCCTACACACACACCTGCTCAGCCCCAC 311
 QY 1005 CTACGGCGTTGGTCCATGATGCTTTGGGCCCTTGACCGATGCCAAGACTAGAGACCA 1064
 DB 310 CTACGGCGTTGGTCCATGATGCTTTGGGCCCTTGACCGATGCCAAGACTAGAGACCA 251
 QY 1065 TGCTGATGATGAGGCTCTGCTTTCTTCTTATGACAGGCTAGATATATACCAAGGGGAGTA 1124
 DB 250 TGCTGATGATGAGGCTCTGCTTTCTTCTTATGACAGGCTAGATATATACCAAGGGGAGTA 191
 QY 1125 CAACCGTTTGGCTCCATTAATGATTAACCACTTAACAAACAA 1170
 DB 190 CAACCGTTTGGCTCCATTAATGATTAACCACTTAACAAACAA 145

RESULT 8
 AA818804/c
 LOCUS
 DEFINITION
 AA818804 538 bp mRNA linear EST 03-JUL-1999
 UI-R-A0-ar-b-11-0-UI s1 UI-R-A0 Rattus norvegicus cDNA clone
 UI-R-A0-ar-b-11-0-UI 3, similar to gi1285286|gb|AA773715|AA773715
 af81c05.r1 Soares Nhhmpu SI Homo sapiens cDNA clone 1048424 5',
 mRNA sequence.
 AA818804
 AA818804.1 GI:4227480
 EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 538)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT On Feb 17, 1998 this sequence version replaced gi:2888390.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Brain library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LNU (info@image.llnl.gov). IMAGE ID=1767889 The following repetitive elements were found in this cDNA sequence:
 105-154, >(CAA)n#Simple_repeat
 Seg primer: M13 Forward
 POLYA=No.

FEATURES
 source Location/Qualifiers
 1..538
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-A0-ar-b-11-0-UI"
 /clone_1lb="UI-R-A0"
 /dev_stage="adult"
 /note="Vector: pTZ19D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 146 a 115 c 102 g 175 t
 ORIGIN
 Query Match 12.5%; Score 168; DB 9; Length 538;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 268; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1023 GAATGCTTTTGGGCGCTTACCGAGTCCCAAGACTAGAGCCATGCTGATGATGAGGCTCT 1082
 DB 284 GAATGCTTTTGGGCGCTTACCGAGTCCCAAGACTAGAGCCATGCTGATGATGAGGCTCT 225
 QY 1083 CGTTCTTCTTATGACAGGCTAGTATATACCAAGGGGAGTCAACCGTTTGGCTGCATA 1142
 DB 224 CGTTCTTCTTATGACAGGCTAGTATATACCAAGGGGAGTCAACCGTTTGGCTGCATA 165

VERSION A1843387.1 GI:5477600
 EST.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 379)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mst@mail.nih.gov

The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the normalized hippocampus library cDNA library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 117-163, >(CAAA)n\$imple_repeat
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..379
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-A01-adz-f-10-0-UI"
 /clone_1db="NIH BMAP_MHI_N"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH BMAP MHI library is a normalized library constructed from mouse hippocampus. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
 TAG_LIB=NIH BMAP_MHI_N
 TAG_TISSUE=hippocampus
 TAG_SEQ=TTGCA"

BASE COUNT 82 a 75 c 92 g 130 t
 ORIGIN

Query Match 13.7%; Score 183; DB 9; Length 379;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 233; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

937 TACAGTGAAGTTACGAGAGTTATGTCGACCCCTACACACACACTGCTCCA 996
 Db 379 TACAGTGAAGTTACGAGAGTTATGTCGACCCCTACACACACACTGCTCCA 320
 997 GCGCCACCTACGCGGTGGTGCATGAATGCTTTTGGCCCTTGACCGATCCCAAGT 1056
 Db 319 GCGCCACCTACGCGGTGGTGCATGAATGCTTTTGGCCCTTGACCGATCCCAAGT 260
 1057 AGGAGCAGTGAAGTGGGTCTGCTTCTTCTTCAATGAGGCTGATATACCA 1116
 Db 259 AGGAGCAGTGAAGTGGGTCTGCTTCTTCTTCAATGAGGCTGATATACCA 200

Qy 1117 GGGGATACAAACCGTTTGTCTCATATTAATGATTAACCATTAACCAACAA 1170
 Db 199 GGGGATACAAACCGTTTGTCTCATATTAATGATTAACCATTAACCAACAA 146

RESULT 6
 LOCUS A1175239/c 266 bp mRNA linear EST 20-JAN-1999
 DEFINITION EST218774 Normalized rat muscle, Bento Soares Rattus sp. cDNA clone
 RMUCER0.3' end, mRNA sequence.
 ACCESSION A1175239
 VERSION A1175239.1 GI:3725877
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 266)
 AUTHORS Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J., Karlavage,A.R. and Adams,M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@igr.org
 Seq primer: M13-21.

FEATURES

source

Location/Qualifiers
 1..266
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2032297"
 /db_xref="taxon:10118"
 /clone="RMUCER0"
 /clone_1db="Normalized rat muscle, Bento Soares"
 /note="Organ: muscle; Vector: pRT3D-Pac; Site 1: EcoRI; Site 2: NotI"

BASE COUNT 65 a 51 c 72 g 78 t
 ORIGIN
 Query Match 13.1%; Score 175; DB 9; Length 266;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 945 CAGTTACGAGAGTTATGCTGCGACCCCTACACACACACTGCTCCAGCCCCAC 1004
 Db 246 CAGTTACGAGAGTTATGCTGCGACCCCTACACACACACTGCTCCAGCCCCAC 187
 Qy 1005 CTAGGCGCTTGTCGATGAATGCTTTTGGCCCTTGACCGATCGCAAGCTAGAGCCA 1064
 Db 186 CTAGGCGCTTGTCGATGAATGCTTTTGGCCCTTGACCGATCGCAAGCTAGAGCCA 127
 Qy 1065 TGCTGATGATGGGTCTGCTTCTTCTTCAATGAGGCTAGTATACCAAGGGGATA 1124
 Db 126 TGCTGATGATGGGTCTGCTTCTTCTTCAATGAGGCTAGTATACCAAGGGGATA 67
 Qy 1125 CAACGTTTGTCTCATATTAATGATTAACCATTAACCAACAA 1170
 Db 66 CAACGTTTGTCTCATATTAATGATTAACCATTAACCAACAA 21

RESULT 7
 LOCUS A1850237 430 bp mRNA linear EST 15-JUL-1999
 DEFINITION UI-M-BG1-a1e-e-06-0-UI.s1 NIH BMAP MSC N Mus musculus cDNA clone
 UI-M-BG1-a1e-e-06-0-UI 3', mRNA sequence.
 ACCESSION A1850237
 VERSION A1850237.1 GI:5494143
 KEYWORDS EST.
 SOURCE house mouse.


```

/db_xref="ATCC (inhost):2022664"
/db_xref="taxon:10118"
/clone="RBRBK70"
/clone_lib="Normalized rat brain, Bento Soares"
/notes="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT      146 a 168 c 128 g 100 t

ORIGIN
Query Match      25.7%; Score 344; DB 9; Length 542;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 CCGTCTCCGGGACCGCCACAGACAGATGATGCCGCCGCCGACCGCCAGCCCGCAGA 259
DB 199 CCGTCTCCGGGACCGCCACAGACAGATGATGCCGCCGCCGACCGCCAGCCCGCAGA 258
QY 260 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCCACA 319
DB 259 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCCACA 318
QY 320 TCCCTTCGGGTCGGGATCCAGACTCCGCAATGTTGGCCAAATTTGGTAAATAT 379
DB 319 TCCCTTCGGGTCGGGATCCAGACTCCGCAATGTTGGCCAAATTTGGTAAATAT 378
QY 380 TAGATGTTGAATTAATTTTAATAGACGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 439
DB 379 TAGATGTTGAATTAATTTTAATAGACGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 438
QY 440 AAAATAGTCCGATCCGACAGGGCGAGAGAGAAATTCGACGTAACCTGTAGAGGGCC 499
DB 439 AAAATAGTCCGATCCGACAGGGCGAGAGAGAAATTCGACGTAACCTGTAGAGGGCC 498
QY 500 GTAATATGAGGTTAATATGAGACAGACGCGTATGACTAAT 543
DB 499 GTAATATGAGGTTAATATGAGACAGACGCGTATGACTAAT 542

RESULT 2
BF420112      494 bp      mRNA      linear      EST 28-NOV-2000
LOCUS         BF420112
DEFINITION    UI-R-BJ2-bpJ-a-08-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone
ACCESSION     BF420112
VERSION       BF420112
KEYWORDS      BF420112.1 GI:11408101
SOURCE        EST.
ORGANISM      Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 494)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  JOURNAL      Genome Res. 6 (9), 791-806 (1996)
  MEDLINE     97044477
  COMMENT     Contact: Soares, MB
              Program for Rat Gene Discovery and Mapping
              University of Iowa
              451 Eckstein Medical Research Building Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: msoares@blue.weeg.uiowa.edu
              Oligo-dT track not found, Not I site shown in beginning of sequence
              is likely internal to the message. cDNA Library Preparation: M.B.
              Soares Lab Clone distribution: clones will be available through
              Research Genetics (www.resgen.com)
              Seq primer: M13 Forward
              POLY-A=No

FEATURES
  source      Location/Qualifiers
              1..494
              /organism="Rattus norvegicus"

```

```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bpJ-a-08-0-UI"
/clone_lib="UI-R-BJ2"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2
library is a subtracted library derived from the following
tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,
atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15
dpc. AV canal at 15 dpc. For a detailed description of
the library from which this clone was derived, please
visit our web site at ratest.eng.uiowa.edu. The
subtraction has been previously described in (Bonaldo,
Lennon and Soares, Genome Research 6:791-806, 1996)
TAG SEQ=None found"

BASE COUNT      128 a 162 c 115 g 89 t

ORIGIN
Query Match      21.6%; Score 289; DB 12; Length 494;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 CCGTCTCCGGGACCGCCACAGACAGATGATGCCGCCGCCGACCGCCAGCCCGCAGA 259
DB 206 CCGTCTCCGGGACCGCCACAGACAGATGATGCCGCCGCCGACCGCCAGCCCGCAGA 265
QY 260 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCCACA 319
DB 266 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCCACA 325
QY 320 TCCCTTCGGGTCGGGATCCAGACTCCGCAATGTTGGCCAAATTTGGTAAATAT 379
DB 326 TCCCTTCGGGTCGGGATCCAGACTCCGCAATGTTGGCCAAATTTGGTAAATAT 385
QY 380 TAGATGTTGAATTAATTTTAATAGACGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 439
DB 386 TAGATGTTGAATTAATTTTAATAGACGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 445
QY 440 AAAATAGTCCGATCCGACAGGGCGAGAGAGAAATTCGACGTAACCT 488
DB 446 AAAATAGTCCGATCCGACAGGGCGAGAGAGAAATTCGACGTAACCT 494

RESULT 3
AA859573      460 bp      mRNA      linear      EST 03-JUL-1999
LOCUS         AA859573
DEFINITION    UI-R-E0-DV-c-04-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
ACCESSION     AA859573
VERSION       AA859573.1 GI:4230119
KEYWORDS      AA859573
SOURCE        EST.
ORGANISM      Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 460)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  JOURNAL      Genome Res. 6 (9), 791-806 (1996)
  MEDLINE     97044477
  COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949093.
              Contact: Soares, MB
              Program for Rat Gene Discovery and Mapping
              University of Iowa
              451 Eckstein Medical Research Building Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: msoares@blue.weeg.uiowa.edu

```

GenCore version 5.1.4_dp5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 15:59:00 ; Search time 1885.21 Seconds
(without alignments)
11511.672 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340
Sequence: 1 ggcggccggcccgacacaa.....ggccgaagggttcgtcgtaga 1340

Scoring table: OLIGO_NUC
Gapco 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_gesba:*
2: em_gesbun:*
3: em_geslin:*
4: em_geslun:*
5: em_gesrov:*
6: em_gesrpl:*
7: em_gestro:*
8: em_hcc:*
9: gb_esc1:*
10: gb_esc2:*
11: gb_hcc:*
12: gb_esc3:*
13: gb_esc4:*
14: gb_esc5:*
15: em_escfun:*
16: em_escrom:*
17: gb_gesb:*
18: em_gesb_hum:*
19: em_gesb_inv:*
20: em_gesb_pln:*
21: em_gesb_vrc:*
22: em_gesb_fun:*
23: em_gesb_man:*
24: em_gesb_mus:*
25: em_gesb_other:*
26: em_gesb_pro:*
27: em_gesb_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	344	25.7	542	9	A1101278
2	289	21.6	494	12	BF420112
3	256	19.1	460	9	AA859573
4	183	15.0	324	9	A1501505
5	183	13.7	379	9	A1843387
6	175	13.1	266	9	A1175239

C	7	175	13.1	430	9	A1850237	A1850237	UT-M-BG1-
C	8	168	12.5	538	9	AA818804	AA818804	UT-R-A0-a
C	9	160	11.9	533	12	BF549922	BF549922	UT-R-E0-b
C	10	157	11.7	361	9	A1835011	A1835011	UI-M-AM1-
C	11	146	10.9	425	10	BB750925	BB750925	BB750925
C	12	140	10.4	395	9	A1717457	A1717457	UT-R-Y0-a
C	13	97	7.2	600	13	B1888557	B1888557	4023-45 M
C	14	80	6.0	408	12	BG795922	BG795922	UTSW-SM38
C	15	70	5.2	228	14	Z28900	Z28900	HSB87F031 S
C	16	70	5.2	337	12	BG186689	BG186689	RST5664 A
C	17	70	5.2	363	14	D60520	D60520	HUM11AE06A
C	18	70	5.2	367	9	A1095813	A1095813	CD20911.x
C	19	70	5.2	383	9	A1189407	A1189407	q401R05.x
C	20	70	5.2	391	9	A1186273	A1186273	q420B08.x
C	21	70	5.2	392	9	AA975235	AA975235	OC36C08.s
C	22	70	5.2	420	9	A1655094	A1655094	WB761A10.x
C	23	70	5.2	427	12	BF223478	BF223478	7G33601.x
C	24	70	5.2	439	9	A1968176	A1968176	W114A06.x
C	25	70	5.2	439	10	AM889795	AM889795	hg22A04.x
C	26	70	5.2	439	10	BE501380	BE501380	7A41H09.x
C	27	70	5.2	443	9	AA773715	AA773715	at81C05.r
C	28	70	5.2	449	9	A1374891	A1374891	ta60G12.x
C	29	70	5.2	456	9	A1917298	A1917298	ts96G09.x
C	30	70	5.2	459	9	A1244212	A1244212	q386B11.x
C	31	70	5.2	477	10	AM197589	AM197589	Xm44G02.x
C	32	70	5.2	480	9	A1401040	A1401040	lt27a12.x
C	33	70	5.2	495	9	A1656926	A1656926	lt48E08.x
C	34	70	5.2	607	9	AU145000	AU145000	AU145000
C	35	70	5.2	773	10	AV729198	AV729198	AV729198
C	36	70	5.2	774	10	AV729057	AV729057	AV729057
C	37	70	5.2	4039	11	BC026312	BC026312	Homo. gap1
C	38	66	4.9	283	12	BG185609	BG185609	RST4560 A
C	39	62	4.6	674	12	BG793800	BG793800	UTSW-SM16
C	40	55	4.1	228	10	AV339026	AV339026	AV339026
C	41	55	4.1	228	10	AV349493	AV349493	AV349493
C	42	55	4.1	226	9	AV253203	AV253203	AV253203
C	43	55	4.1	274	10	BB197021	BB197021	BB197021
C	44	55	4.1	277	10	BB431921	BB431921	BB431921
C	45	55	4.1	295	10	BB188777	BB188777	BB188777

ALIGNMENTS

RESULT 1
A1101278
LOCUS 542 bp mRNA linear EST 31-JAN-1999
DEFINITION EST210567 Normalized rat brain, Bentos Soares Rattus sp. cDNA clone
RBRER70.3' end, mRNA sequence.
ACCESSION A1101278
VERSION A1101278.1 GI:3704621
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 542)
AUTHORS Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
1..542
/organism="Rattus sp."

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;

Query Match 2.9%; Score 39; DB 23; Length 2118;
Best Local Similarity 100.0%; Pred. No. 66-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 946 AGTTACGACGACGATTATGCTGCCGACCCCTACCCACAC 984

Db 1300 AGTTACGACGACGATTATGCTGCCGACCCCTACCCACAC 1338

RESULT 14
AAH03678
ID AAH03678 standard; cDNA; 893 BP.

AC AAH03678;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:513.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;

XX MPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -

XX Claim 1; SEQ ID 513; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH1632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 893 BP; 225 A; 228 C; 236 G; 199 T; 5 other;

Query Match 2.6%; Score 35; DB 22; Length 893;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 415 AAGGATTGCTTCGTAACCTTCGAAATAGTGC 449

Db 686 AAGGATTGCTTCGTAACCTTCGAAATAGTGC 720

RESULT 15
AAS31173/C
ID AAS31173 standard; cDNA; 1295 BP.

XX AAS31173;

XX 04-DEC-2001 (first entry)

XX Human diagnostic and therapeutic polynucleotide (DITHP) #48.

XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
XX cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
XX acquired immune deficiency syndrome; AIDS; autoimmune disorder;
XX respiratory disorder; ss.

XX OS

XX Homo sapiens.

XX WO200162927-A2.

XX 30-AUG-2001.

XX 21-FEB-2001; 2001WO-US06059.

XX 24-FEB-2000; 2000US-0184693.

XX 24-FEB-2000; 2000US-0184697.

XX 24-FEB-2000; 2000US-0184698.

XX 24-FEB-2000; 2000US-0184768.

XX 24-FEB-2000; 2000US-0184769.

XX 24-FEB-2000; 2000US-0184770.

XX 24-FEB-2000; 2000US-0184771.

XX 24-FEB-2000; 2000US-0184772.

XX 24-FEB-2000; 2000US-0184773.

XX 24-FEB-2000; 2000US-0184774.

XX 24-FEB-2000; 2000US-0184776.

XX 24-FEB-2000; 2000US-0184777.

XX 24-FEB-2000; 2000US-0184797.

XX 24-FEB-2000; 2000US-0184813.

XX 24-FEB-2000; 2000US-0184837.

XX 24-FEB-2000; 2000US-0184841.

XX 24-FEB-2000; 2000US-0185213.

XX 24-FEB-2000; 2000US-0185216.

XX 12-MAY-2000; 2000US-0203785.

XX 15-MAY-2000; 2000US-0204226.

XX 16-MAY-2000; 2000US-0204525.

XX 16-MAY-2000; 2000US-0204821.

XX 16-MAY-2000; 2000US-0204908.

XX 16-MAY-2000; 2000US-0205232.

XX 17-MAY-2000; 2000US-0204815.

XX 17-MAY-2000; 2000US-0204863.

XX 17-MAY-2000; 2000US-0205221.

XX 17-MAY-2000; 2000US-0205285.

XX 17-MAY-2000; 2000US-0205286.

XX 17-MAY-2000; 2000US-0205287.

XX 17-MAY-2000; 2000US-0205323.

CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.
XX
SQ Sequence 3189 BP; 1005 A; 554 C; 631 G; 999 T; 0 other;
Query Match 3.5%; Score 47; DB 24; Length 3189;
Best Local Similarity 100.0%; Pred. No. 4.1e-09;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1068 TGATGATGTGGGTCTCGTCTTTCTTCTTCATTCAGGCTAGTATATACC 1114
DB 515 TGATGATGTGGGTCTCGTCTTTCTTCTTCATTCAGGCTAGTATATACC 561
RESULT 12
AAK53229/c
ID AAK53229 standard; cDNA; 1164 BP.
XX
AC AAK53229;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2758.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSEQ-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZM;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
XX
PT P-PSDB; AAM80096.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 4968; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;
Query Match 3.2%; Score 43; DB 22; Length 1164;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 634 CTGTTGCCAGGCCAACAGAGGATCTTCCATGTACAGTG 676
DB 173 CTGTTGCCAGGCCAACAGAGGATCTTCCATGTACAGTG 131
RESULT 13
AAS93634
ID AAS93634 standard; cDNA; 2118 BP.
XX
AC AAS93634;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29438.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSEQ-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
PT P-PSDB; ABG29447.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 29438; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptides and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

XX Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;
 SQ Query Match 5.2%; Score 70; DB 21; Length 2372;
 Best Local Similarity 100.0%; Pred. No. 4e-18;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGAGCCATGCTGATGATGCTGCTCTCTCTTTCTTCATTCAGGCT 1104
 DB 2133 GATGCCAAGACTAGAGCCATGCTGATGATGCTGCTCTCTCTTTCTTCATTCAGGCT 2192

QY 1105 AGTATATACC 1114
 DB 2193 AGTATATACC 2202

RESULT 10
 ABN34672
 ID ABN34672 standard; DNA; 60 BP.
 AC ABN34672;
 XX
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human spliced transcript detection oligonucleotide SEQ ID NO:7420.
 KW Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200210449-A2.
 PD 07-FEB-2002.
 PF 20-JUL-2001; 2001WO-1B01903.
 PR 28-JUL-2000; 2000US-221607P.
 PR 02-MAY-2001; 2001US-287724P.
 XX
 PA (COMP-) COMPUGEN INC.
 XX
 PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 XX
 DR WPI; 2002-257363/30.
 XX
 PT New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes -
 XX
 PS Example 1; SEQ ID 7420; 47pp; English.
 XX
 CC The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridising selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterising the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in

CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX

SQ Sequence 60 BP; 11 A; 14 C; 16 G; 19 T; 0 other;
 Query Match 3.6%; Score 48; DB 24; Length 60;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGAGCCATGCTGATGATGCTGCTCTCTTCT 1092
 DB 13 GATGCCAAGACTAGAGCCATGCTGATGATGCTGCTCTCTTCT 60

RESULT 11
 ABR34602
 ID ABR34602 standard; cDNA; 3189 BP.
 AC ABR34602;
 XX
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human cDNA for novel secreted protein, SEQ ID 371.
 XX
 KW Human; ss; gene; secreted protein; immune deficiency; viral infection;
 KW bacterial infection; fungal infection; autoimmune disorder; burn;
 KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
 KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
 KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
 KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
 KW lymphoid cell deficiency.
 XX
 OS Homo sapiens.
 XX
 PN WO200177290-A2.
 PD 18-OCT-2001.
 PF 29-MAR-2001; 2001WO-US10295.
 PR 06-APR-2000; 2000US-194941P.
 XX
 PA (GENY) GENETICS INST INC.
 XX
 PI Wong GG, Clark HF, Fechtel K, Agostino MT, Howes SH, Reenick RJ;
 PI Gulukota K, Graham JR;
 XX
 DR WPI; 2002-179323/23.
 XX
 PT Six hundred and twenty five polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders
 PT -
 XX
 PS Claim 1; Page 188-189; 339pp; English.
 XX
 CC The invention relates to 623 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins, their complements and sequences that hybridise to them.
 CC Also included are a vector comprising the polynucleotide, a host cell
 CC transformed with the vector, the proteins encoded by the
 CC polynucleotides, antibodies that bind to the proteins and identification
 CC of modulators of the proteins or the expression of the polynucleotide.
 CC The polynucleotides can be used as probes for the identification
 CC and isolation of full length cDNA and genomic DNA. The polynucleotides
 CC and proteins can also be used as nutritional supplements. The protein
 CC is useful in the treatment of various immune deficiencies and disorders
 CC such as viral infections, bacterial infections, fungal infections,
 CC autoimmune disorders (e.g. Rheumatoid arthritis, multiple sclerosis,
 CC autoimmune thyroiditis and diabetes) and allergic reactions and

XX AAK52245;
 AC ID
 DT 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 790.
 XX
 DE Human polynucleotide SEQ ID NO 790.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 DR WPI; 2001-476283/51.
 DR P-PSDB; AAM79112.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 2643-2645; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
 Query Match 5.2%; Score 70; DB 22; Length 1800;
 Best Local Similarity 100.0%; Pred. No. 4.2e-18;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1045 GATGCCAAGACTAGAGACCATGCTGATGATGCGGCTCTGTTCTTTCATTGCAAGCT 1104
 DB 1600 GATGCCAAGACTAGAGACCATGCTGATGATGCGGCTCTGTTCTTTCATTGCAAGCT 1659
 QY 1105 AGTATATACC 1114
 DB 1660 AGTATATACC 1669

RESULT 9
 AAA07075
 ID AAA07075 standard; cDNA; 2372 BP.
 XX
 AC AAA07075;
 XX
 DT 03-JUL-2000 (first entry)
 XX
 DE cDNA encoding human ataxin-2 binding protein (A2BP).
 XX
 KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
 KW neuronal plasticity; cellular degeneration signal transduction pathway;
 KW selective RNA transport; spinocerebellar ataxia type-2;
 KW hyperproliferative disorder; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 987..1979
 FT CDS /*tag= a
 FT /product= "Human A2BP"
 XX
 PN MO200012710-A1.
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20156.
 XX
 PR 01-SEP-1998; 98US-0145391.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Pulst SM, Shibata H;
 PI WPI; 2000-237873/20.
 DR P-PSDB; AAY81462.
 XX
 PT Nucleic acids encoding an ataxin-2 binding protein useful for
 PT inhibiting the expression of active proteins from the SCA2 gene for the
 PT treatment of spinocerebellar ataxia type-2 -
 XX
 PS Claim 6; Page 74-77; 82pp; English.
 XX
 CC This sequence represents cDNA encoding human ataxin-2 binding protein
 CC (A2BP). Nucleotide sequences encoding human A2BP were originally
 CC isolated in an adult brain cDNA library using the yeast two hybrid
 CC method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
 CC function that is encoded by the SCA2 gene located on chromosome 12. SCA2
 CC has been linked to the autosomal dominant neurodegenerative disorder
 CC spinocerebellar ataxia type-2. Individuals afflicted with the disease
 CC exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
 CC containing a polyglutamine stretch of about 35-39 residues, whereas that
 CC of normal individuals contains approximately 22 contiguous glutamine
 CC residues. A2BP and ataxin-2 are components of a cellular degeneration
 CC signal transduction pathway. The pathogenic expanded form of ataxin-2 has
 CC a higher affinity for A2BP relative to normal ataxin-2; the presence of
 CC the expanded form is likely to promote degeneration. A2BP and ataxin have
 CC also been found to have a role in gene regulation. The binding of A2BP to
 CC ataxin-2 plays an important role in controlling gene expression via the
 CC targeting of transport of specific RNAs, selective RNA transport being
 CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
 CC early in embryonic development. Both ataxin-2 and A2BP are able to bind
 CC RNA, and are essential components of the RNA localisation network that
 CC establishes cellular polarity in embryogenesis. In highly differentiated,
 CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
 CC function and are required for neuronal plasticity. A2BP nucleic acids may
 CC be used for the recombinant production of A2BP proteins or fragments
 CC thereof according to standard methodologies. For example, an A2BP protein
 CC with an ataxin-2 or RNA binding capability but no signal transduction
 CC function can be used as a dominant negative inhibitor of the cellular
 CC degeneration signal transduction pathway. A2BP proteins with a signal
 CC transduction function can be used to treat hyperproliferative disorders
 CC (e.g., cancer) via stimulation of the cellular degeneration pathway.

XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 XX 11-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Oca T, Isegai T, Nishikawa T, Hayaehi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 XX Claim 3; SEQ ID 6040; 2537pp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesizing 5602
 XX full-length cDNAs defined in the specification. Where a primer set
 XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 XX to the complementary strand of a polynucleotide which comprises one of
 XX the 5602 nucleotide sequences defined in the specification, where the
 XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 XX of an oligonucleotide comprising a sequence complementary to the
 XX complementary strand of a polynucleotide which comprises a 5'-end
 XX sequence and an oligonucleotide comprising a sequence complementary to a
 XX polynucleotide which comprises a 3'-end sequence, where the
 XX oligonucleotide comprises at least 15 nucleotides and the combination of
 XX the 5'-end sequence/3'-end sequence is selected from those defined in
 XX the specification. The primer sets can be used in antisense therapy and
 XX in gene therapy. The primers are useful for synthesizing polynucleotides,
 XX particularly full-length cDNAs. The primers are also useful for the
 XX detection and/or diagnosis of the abnormality of the proteins encoded by
 XX the full-length cDNAs. The primers allow obtaining of the full-length
 XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 XX represent oligonucleotides, all of which are used in the exemplification
 XX of the present invention.
 XX
 XX Sequence 607 BP; 140 A; 152 C; 179 G; 129 T; 7 other;
 XX
 XX Query Match 5.2%; Score 70; DB 22; Length 607;
 XX Best Local Similarity 100.0%; Pred. No. 4.9e-18;
 XX Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1045 GATGCCAAGACTAGAGCCATGCTGATGATGAGGCTCTGTTCTTTCATTGACAGGCT 1104
 XX DB 201 GATGCCAAGACTAGAGCCATGCTGATGATGAGGCTCTGTTCTTTCATTGACAGGCT 142
 XX
 XX QY 1105 AGTATATACC 1114
 XX DB 141 AGTATATACC 132
 XX
 XX RESULT 7
 XX AAH13824
 XX ID AAH13824 standard; cDNA; 1513 BP.
 XX AC AAH13824;
 XX XX
 XX 26-JUN-2001 (first entry)
 XX XX

DE Human cDNA sequence SEQ ID NO:10786.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 XX 11-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Oca T, Isegai T, Nishikawa T, Hayaehi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 XX Claim 8; SEQ ID 10786; 2537pp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesizing 5602
 XX full-length cDNAs defined in the specification. Where a primer set
 XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 XX to the complementary strand of a polynucleotide which comprises one of
 XX the 5602 nucleotide sequences defined in the specification, where the
 XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 XX of an oligonucleotide comprising a sequence complementary to the
 XX complementary strand of a polynucleotide which comprises a 5'-end
 XX sequence and an oligonucleotide comprising a sequence complementary to a
 XX polynucleotide which comprises a 3'-end sequence, where the
 XX oligonucleotide comprises at least 15 nucleotides and the combination of
 XX the 5'-end sequence/3'-end sequence is selected from those defined in
 XX the specification. The primer sets can be used in antisense therapy and
 XX in gene therapy. The primers are useful for synthesizing polynucleotides,
 XX particularly full-length cDNAs. The primers are also useful for the
 XX detection and/or diagnosis of the abnormality of the proteins encoded by
 XX the full-length cDNAs. The primers allow obtaining of the full-length
 XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 XX represent oligonucleotides, all of which are used in the exemplification
 XX of the present invention.
 XX
 XX Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;
 XX
 XX Query Match 5.2%; Score 70; DB 22; Length 1513;
 XX Best Local Similarity 100.0%; Pred. No. 4.3e-18;
 XX Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1045 GATGCCAAGACTAGAGCCATGCTGATGATGAGGCTCTGTTCTTTCATTGACAGGCT 1104
 XX DB 1313 GATGCCAAGACTAGAGCCATGCTGATGATGAGGCTCTGTTCTTTCATTGACAGGCT 1372
 XX
 XX QY 1105 AGTATATACC 1114
 XX DB 1373 AGTATATACC 1382
 XX
 XX RESULT 8
 XX AAK52245
 XX ID AAK52245 standard; cDNA; 1800 BP.
 XX

PA (SCIO-) SCIOS INC.
 XX Stanon LM, White RT;
 XX WPI, 2002-010779/01.
 DR P-PSDB; AAU70146.
 XX Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX
 PS Claim 1; Fig 1; 189pp; English.
 XX
 CC The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 XX
 SQ Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;
 Query Match 100.0%; Score 1340; DB 24; Length 1340;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 541 AATAAAAAGCCGTGAACCCCTACACCAATGCTGGAATTAATCAAGTTGGCCCG 600
 Qy GTCTACAGCCCGGACTTCTATGACAGACCGGTGTTGGCCAGGCCAACCAGAGGGA 660
 Db 601 GTCTACAGCCCGGACTTCTATGACAGACCGGTGTTGGCCAGGCCAACCAGAGGGA 660
 Qy 661 TCTTCATGTCAGTGGCCCGGACTTCTATGACAGACCGGTGTTGGCCAGGCCAACCAGAGGGA 720
 Db 661 TCTTCATGTCAGTGGCCCGGACTTCTATGACAGACCGGTGTTGGCCAGGCCAACCAGAGGGA 720
 Qy 721 TATCCGCGCCGACTGCTGACGTCGATACCGAGGGCTCACTTGCAGCGCGGTGTCG 780
 Db 721 TATCCGCGCCGACTGCTGACGTCGATACCGAGGGCTCACTTGCAGCGCGGTGTCG 780
 Qy 781 ACCGTCATCAACCTTCTAGAGTGCAGCGCCCGCCACCCCAATCCGGCTATGCGGA 840
 Db 781 ACCGTCATCAACCTTCTAGAGTGCAGCGCCCGCCACCCCAATCCGGCTATGCGGA 840
 Qy 841 GTAGTGTATCAAGAGCCAGTGTATGGAATTAATGCTACAGGGTGTACGCTGCATAC 900
 Db 841 GTAGTGTATCAAGAGCCAGTGTATGGAATTAATGCTACAGGGTGTACGCTGCATAC 900
 Qy 901 CGCTACGCCCGCCAGCCCGCTGACAGTGTCTGCTGCTACAGTGAAGTACGAGGAGTT 960
 Db 901 CGCTACGCCCGCCAGCCCGCTGACAGTGTCTGCTGCTACAGTGAAGTACGAGGAGTT 960
 Qy 961 TATGTCGCCGACCCCTACACCAACACTTGTCTGACGCCCGCCACTACAGGCGTGTGTC 1020
 Db 961 TATGTCGCCGACCCCTACACCAACACTTGTCTGACGCCCGCCACTACAGGCGTGTGTC 1020
 Qy 1021 ATGAATGCTTTTGGCCGCTTGGACCGATGCCAAGACTAGAGCCATGCTGATGATGCGGT 1080
 Db 1021 ATGAATGCTTTTGGCCGCTTGGACCGATGCCAAGACTAGAGCCATGCTGATGATGCGGT 1080
 Qy 1081 CTGCTTCTTTTCTTATGACAGCTAGTATATACCAAGGGGATACCAACGTTTGTCTCA 1140
 Db 1081 CTGCTTCTTTTCTTATGACAGCTAGTATATACCAAGGGGATACCAACGTTTGTCTCA 1140
 Qy 1141 TATTAATGATTAACCACTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1200
 Db 1141 TATTAATGATTAACCACTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1200
 Qy 1201 ACCTTCAATGTGGGAGAGAAAGCTTTCGAGGCCGAGTGTGCGACATGCAGT 1260
 Db 1201 ACCTTCAATGTGGGAGAGAAAGCTTTCGAGGCCGAGTGTGCGACATGCAGT 1260
 Qy 1261 AGGACATCACTTTAGCAACTCAAAAGAAACAAAGAAAAAATTAAGC 1320
 Db 1261 AGGACATCACTTTAGCAACTCAAAAGAAACAAAGAAAAAATTAAGC 1320
 Qy 1321 GGCCGAAGGGGTGCTGAGA 1340
 Db 1321 GGCCGAAGGGGTGCTGAGA 1340

RESULT 2
 AAKS4001/C
 ID AAKS4001 standard; cDNA; 327 BP.
 XX AAKS4001;
 AC 16-NOV-2001 (first entry)
 DT
 XX
 DE Murine transcription associated protein encoding cDNA SEQ ID 566.
 XX Murine; liver; gene library; amino acid synthesis; binding protein;
 KW cell metabolism; energy metabolism; fatty acid metabolism; synthesis;
 KW phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;
 KW replication; transcription; translation; transport protein; ss.
 OS Mus musculus.
 XX
 PN DE20103510-UI.

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:52:03 ; Search time 313.615 Seconds
(without alignments)
9625.317 Million cell updates/sec

Title: US-09-809-545A-1
Perfect score: 1340
Sequence: 1 gcggcgcccccgcacacat.....ggccgaagggtcgtccta 1340

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1340	100.0	1340	24	AAS94693	Rat secreted factor
2	91	6.8	327	22	AAK54001	Murine transcript
3	70	5.2	406	23	AAS93632	DNA encoding novel
4	70	5.2	481	23	AAS93633	DNA encoding novel
5	70	5.2	539	24	ABK34530	Human CDNA clone
6	70	5.2	607	22	AAH09205	Human CDNA clone
7	70	5.2	1513	22	AAH13824	Human CDNA sequenc
8	70	5.2	1800	22	AAK52245	Human polynucleoti
9	70	5.2	2372	21	AAA07075	CDNA encoding huma

10	48	3.6	60	24	ABN34672	Human spliced tran
11	47	3.5	3189	24	ABK34602	Human CDNA for nov
12	43	3.2	1164	22	AAK53329	Human polynucleoti
13	39	2.9	2118	23	AAS93634	DNA encoding novel
14	35	2.6	893	22	AAH03678	Human CDNA clone
15	35	2.6	1295	22	AAS31173	Human diagnostic a
16	32	2.4	125	22	AAK20380	Human brain expres
17	32	2.4	125	22	AAK46486	Human bone marrow
18	32	2.4	125	22	AAI52354	Probe #21040 used
19	32	2.4	125	22	ABK20830	Human genome-deriv
20	32	2.4	394	20	AAV87007	EST clone AY147.
21	32	2.4	473	22	AAK07716	Human brain expres
22	32	2.4	473	22	AAK33549	Human bone marrow
23	32	2.4	473	22	AAI39304	Probe #7990 used t
24	32	2.4	473	22	ABK08416	Human genome-deriv
25	32	2.4	493	23	ABV40823	Human prostate exp
26	32	2.4	511	23	ABV38877	Human prostate exp
27	30	2.2	979	24	ABK29216	Oligonucleotide fo
28	30	2.2	979	24	ABK29217	Oligonucleotide fo
29	30	2.2	1168	24	ABK47032	Oligonucleotide fo
30	30	2.2	1168	24	ABK47033	Oligonucleotide fo
31	30	2.2	1574	21	AAK18023	Lung cancer associ
32	30	2.2	2308	24	ABK39167	Oligonucleotide fo
33	30	2.2	2308	24	ABK39167	Oligonucleotide fo
34	30	2.2	2365	24	ABK17068	Oligonucleotide fo
35	30	2.2	2365	24	ABK17069	Oligonucleotide fo
36	30	2.2	2395	24	ABK13720	Oligonucleotide fo
37	30	2.2	2395	24	ABK13721	Oligonucleotide fo
38	30	2.2	5695	22	AAS46538	Tumour suppressor
39	30	2.2	5898	22	AAS45471	Chemically pretrea
40	30	2.2	5898	24	ABK33940	Human immune syste
41	30	2.2	5898	24	ABK28329	DNA transcripion
42	30	2.2	6126	22	AAS46574	Tumour suppressor
43	30	2.2	6126	24	ABK33831	Human immune syste
44	30	2.2	6154	24	ABK33305	Human immune syste
45	30	2.2	6510	22	AAS46647	Tumour suppressor

ALIGNMENTS

RESULT 1	AAS94693	Rat secreted factor DNA clone P0184_D11 #1.
XX	XX	AAK54693 ;
XX	XX	12-MAR-2002 (first entry)
DE	XX	Rat secreted factor DNA clone P0184_D11 #1.
XX	XX	Rat; secreted factor polypeptide; cardiac disease; renal disease; kidney;
KW	XX	inflammatory disease; congestive heart failure; myocarditis; asthma; ss;
KW	XX	dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
KW	XX	myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
KW	XX	atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
KW	XX	renal infarction; hereditary nephritis; polycystic kidney disease;
KW	XX	chronic renal failure; renal vein thrombosis; medullary sponge kidney;
KW	XX	rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;
KW	XX	graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
KW	XX	Alzheimer's disease; gene therapy.
OS	XX	Rattus norvegicus.
PN	XX	W0200174901-A2.
PD	XX	11-OCT-2001.
XX	XX	23-MAR-2001; 2001WO-US09555.
PF	XX	23-MAR-2000; 2000US-193548P.
PR	XX	14-MAR-2001; 2001US-0809545.
XX	XX	

THIS PAGE BLANK (U.S.)

Matches	67;	Conservative	.0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy 881	AGG	TGTTACG	TACG	TACCG	TACCG	CCAGCC	CCAGCC	CCCTG	CCCTG
Db 34697	AGG	TGTTACG	TACG	TACCG	TACCG	CCAGCC	CCAGCC	CCCTG	CCCTG
Qy 941	GTG	ACAG	947						
Db 34757	GTG	ACAG	34763						

Search completed: March 15, 2003, 22:48:10
 Job time : 3992.54 secs

RESULT 15	AC120661	115860 bp	DNA	linear	HTG 23-JUL-2002
LOCUS	Rattus norvegicus clone CH230-34C2, *** SEQUENCING IN PROGRESS ***				
DEFINITION	21 unordered pieces.				
ACCESSION	AC120661.2	GI:21902877			
VERSION	HTG; HTGS PHASE1.				
KEYWORDS	Rattus norvegicus				
SOURCE	Rattus norvegicus				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 115860)				
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochvic,J., Kurehi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lied,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 115860)				
AUTHORS	Worley,K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 115860)				
AUTHORS	Worley,K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
COMMENT	On Jul 18, 2002 this sequence version replaced gi:20514522. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/				
CONTACT	hgsc-help@bcm.tmc.edu				
PROJECT INFORMATION	Center project name: GxLC Center clone name: CH230-34C2 Summary Statistics				
CHEMISTRY	Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 99693 bases at least Q40 Consensus quality: 101936 bases at least Q30 Consensus quality: 102743 bases at least Q20				
NOTE	Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.				
FEATURES	Location/Qualifiers				
SOURCE	1. 115860 /organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="CH230-34C2"				
BASE COUNT	33039 a 22417 c 22786 g 30574 t 7044 others				
ORIGIN					
Query Match	5.0% Score 67 DB 2 Length 115860				
Best Local Similarity	100.0% Pred. No. 1e-28				

```
repeat_region complement(14642. .14755)
repeat_region /rpt family="L1"
repeat_region 17429. .18370
repeat_region /rpt family="MSTAR"
repeat_region complement(19470. .19819)
repeat_region /rpt family="MLT1"
repeat_region complement(20352. .20627)
repeat_region /rpt family="Alu"
repeat_region 23808. .24058
repeat_region /rpt family="Alu"
repeat_region 27654. .28217
repeat_region /rpt family="L1"
repeat_region complement(32764. .32956)
repeat_region /rpt family="L1"
repeat_region complement(33392. .33503)
repeat_region /rpt family="Alu"
repeat_region 34586. .34870
repeat_region /rpt family="Alu"
repeat_region 34936. .35489
repeat_region /rpt family="MLT2B2"
repeat_region complement(36940. .37172)
repeat_region /rpt family="Alu"
repeat_region 37173. .37437
repeat_region /rpt family="THE1"
repeat_region 41428. .41745
repeat_region /rpt family="THE1"
repeat_region complement(42057. .43927)
repeat_region /rpt family="L1"
repeat_region 43924. .44555
repeat_region /rpt family="L1"
repeat_region 45128. .45367
repeat_region /rpt family="Alu"
repeat_region 45467. .45704
repeat_region /rpt family="Alu"
repeat_region 46193. .46578
repeat_region /rpt family="MER25"
repeat_region complement(46779. .47141)
repeat_region /rpt family="THE1"
repeat_region complement(48265. .48550)
repeat_region /rpt family="Alu"
repeat_region 49630. .50221
repeat_region /rpt family="MER41"
repeat_region 50781. .51096
repeat_region /rpt family="Alu"
repeat_region complement(51219. .51439)
repeat_region /rpt family="Alu"
repeat_region 51462. .51708
repeat_region /rpt family="MER33"
repeat_region complement(52042. .52321)
repeat_region /rpt family="Alu"
repeat_region 54001. .54092
repeat_region /rpt family="Alu"
repeat_region complement(55760. .56014)
repeat_region /rpt family="Alu"
repeat_region 56616. .56899
repeat_region /rpt family="Alu"
repeat_region complement(66463. .66747)
repeat_region /rpt family="Alu"
repeat_region complement(66908. .67180)
repeat_region /rpt family="Alu"
repeat_region complement(70389. .70704)
repeat_region /rpt family="MER1"
repeat_region complement(72166. .72421)
repeat_region /rpt family="MER33"
repeat_region 72501. .72769
repeat_region /rpt family="Alu"
repeat_region 72843. .72956
repeat_region /rpt family="MER5"
repeat_region complement(76303. .76564)
repeat_region /rpt family="Alu"
repeat_region 97534. .97700
repeat_region /rpt family="MER3"
repeat_region complement(98469. .98526)
```

```
repeat_region /rpt family="Alu"
repeat_region complement(99089. .99381)
repeat_region /rpt family="Alu"
repeat_region 101648. .101930
repeat_region /rpt family="Alu"
repeat_region 102790. .103056
repeat_region /rpt family="Alu"
repeat_region 104010. .104277
repeat_region /rpt family="Alu"
repeat_region complement(111121. .111263)
repeat_region /rpt family="MERS"
repeat_region complement(112263. .112476)
repeat_region /rpt family="MER30"
repeat_region 112956. .113234
repeat_region /rpt family="Alu"
repeat_region 114385. .114657
repeat_region /rpt family="Alu"
repeat_region complement(117031. .117194)
repeat_region /rpt family="MER20"
repeat_region complement(118464. .118770)
repeat_region /rpt family="Alu"
repeat_region 119271. .119549
repeat_region /rpt family="Alu"
repeat_region complement(122831. .122940)
repeat_region /rpt family="MER41"
repeat_region 123471. .123742
repeat_region /rpt family="Alu"
repeat_region complement(129860. .129972)
repeat_region /rpt family="Alu"
repeat_region 135235. .135310
repeat_region /rpt family="MLT1"
repeat_region 135874. .136157
repeat_region /rpt family="Alu"
repeat_region 139775. .140048
repeat_region /rpt family="Alu"
repeat_region 141734. .141814
repeat_region /rpt family="MIR"
repeat_region 143071. .143352
repeat_region /rpt family="Alu"
repeat_region 148286. .148773
repeat_region /rpt family="MER1"
repeat_region 150431. .150707
repeat_region /rpt family="Alu"
repeat_region 155151. .155387
repeat_region /rpt family="Alu"
repeat_region complement(156101. .156213)
repeat_region /rpt family="MER20"
repeat_region complement(156636. .156920)
repeat_region /rpt family="Alu"
repeat_region complement(159633. .159931)
repeat_region /rpt family="Alu"
repeat_region complement(160378. .160658)
repeat_region /rpt family="Alu"
repeat_region 161617. .162140
repeat_region /rpt family="Alu"
repeat_region complement(163200. .163315)
repeat_region /rpt family="MIR"
BASE COUNT 48114 a 34256 c 33738 g 48430 t
ORIGIN
```

Query Match 5.2%; Score 70; DB 9; Length 164538;
Best Local Similarity 100.0%; Pred.No. 1.5e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGGAGCCATCTCATGTGCGTCTCGTCTCTTCTTCATTGAGGCT 1104
|||||
DB 61328 GATGCCAAGACTAGGAGCCATCTCATGTGCGTCTCGTCTCTTCTTCATTGAGGCT 61269
|||||

QY 1105 AGTATATACC 1114
|||||
DB 61268 AGTATATACC 61259
|||||

```
Db 180 GATGCCAAGACTAGGAGCCATGCTGATGTGGTCTGCTTTCTTCATTGCAGGCT 239
|||||
Qy 1105 AGTATATACC 1114
|||||
Db 240 AGTATATACC 249
|||||

RESULT 13
AF109106 3348 bp mRNA linear PRI 01-MAR-2002
LOCUS Homo sapiens hexaribonucleotide binding protein 1 isoform beta
DEFINITION (HNRBP1) mRNA, complete cds.
ACCESSION AF109106
VERSION AF109106.1 GI:19032367
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3348)
AUTHORS Chen, W., Chu, Z.-L., Blough, R.I., Liu, L., Hoppes, B. and
Winkelman, J.C.
TITLE Molecular Cloning and Chromosomal Localization of a Novel Human
Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene
Homologous to fox-1 in Caenorhabditis elegans
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3348)
AUTHORS Chen, W., Chu, Z.-L., Blough, R.I., Liu, L., Hoppes, B. and
Winkelman, J.C.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1998) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave. ML
0508, Cincinnati, OH 45267-0508, USA
FEATURES
source
1..3348
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
gene 1..3348
/genes="HNRBP1"
CDS 254..1441
/genes="HNRBP1"
/notes="similar"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AAL83406.1"
/db_xref="GI:19032368"
/translations="MLASQGVLLHPYGVPMIVPAAPYLPLGLIQNGEAAAPDTMAQP
YASQAFAPPONGIPAEYTAHPHPAPEYTGTTVPEHTLNLYPPAQTHSPQSDPTSA
OTVSGTATDDAAPTQDQTOPSNTENKSKPKRLHVSNIPEFRDPLROMFGQF
GKILDVILFNERGSKGFGVTFENSADAREKHLGCTVVEGRKLEVNNAARVMTN
KKTVNTYNGKLVNVGVGYSPFYAGTVLLCOANQEGSSMTSAPSSLVYTSMPGF
PYPAATAAAYRGARHLRGRRTVNTFRAAAPPPPIPAYGGVYQEPVGNKLLQGGY
AAVRYAQPTATAAAYSDRNQFVFAADEISCSNTSAVTDFMLPTPTTLLQPPPTA
LVP"
polyA_signal 3299..3304
/gene="HNRBP1"
BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN
Query Match 5.2%; Score 70; DB 9; Length 3348;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1045 GATGCCAAGACTAGGAGCCATGCTGATGTGGTCTGCTTTCTTCATTGCAGGCT 1104
|||||
Db 1462 GATGCCAAGACTAGGAGCCATGCTGATGTGGTCTGCTTTCTTCATTGCAGGCT 1521
|||||
Oy 1105 AGTATATACC 1114
|||||
```

```
Db 1522 AGTATATACC 1531
|||||

RESULT 14
AC005774/c 164538 bp DNA linear PRI 05-OCT-1998
LOCUS Homo sapiens chromosome 16, BAC clone 2603 (LANL), complete
DEFINITION sequence.
ACCESSION AC005774
VERSION AC005774.1 GI:3702264
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164538)
AUTHORS Ricke, D.C.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 164538)
AUTHORS Ricke, D.C., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.
and Deaven, L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 164538)
AUTHORS Ricke, D.C., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.
and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source
1..164538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="2603"
1378..1665
/rpt_family="Alu"
complement(2144..2320)
/rpt_family="MER21"
complement(2198..2416)
/rpt_family="MER39"
complement(2450..2725)
/rpt_family="Alu"
complement(3157..3393)
/rpt_family="Alu"
complement(3524..3792)
/rpt_family="Alu"
5731..6020
/rpt_family="Alu"
complement(7557..7727)
/rpt_family="MER4"
3282..3928
/rpt_family="LTR8"
10055..10310
/rpt_family="THE1"
10472..10618
/rpt_family="L1"
10839..10939
/rpt_family="Alu"
complement(11109..11357)
/rpt_family="Alu"
12050..12312
/rpt_family="Alu"
```



```

TITLE      Direct Submission
JOURNAL    Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
           University of Cincinnati College of Medicine, The Vontz Center for
           Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
FEATURES   Location/Qualifiers
source     1..2000
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /chromosome="16"
           /map="16p13.3"
gene       1..2000
           /genes="HRNBPI"
           /note="HRNBPI"
CDS        256..1512
           /gene="HRNBPI"
           /product="RNA-binding protein"
           /codon_start=1
           /product="hexaribonucleotide binding protein 1 isoform
           gamma"
           /protein_id="AAU71904.1"
           /db_xref="GI:18461367"
           /translation="MLASQGVLLHPYGVMPVPAAPYLPGLIQGNEAAAAADPTWAQP
           YASAOFPAPNGIPAEVTAHPHPAPEYTGOTTVPHEITLNLVPPAOTHSSEQSPADTSA
           OTVSGTATDDDAAPTDGQOPOTSPSTENKSKOPKELHVSNIPIFRDPLDROMFGQF
           GKILDVEIIFNERSKGFVTFENSADADRAREKLGHTVEGRKIEVNATARVMTN
           KKTNYITNGKLNPNVGVAVSPFYAGTIVLLCOAQEGSSMTSAPSSLYITISAMEGF
           PYPAATAAAYRGAAHLRGRGRTVNTFRAAAPPPPIPAYGGVYQEPVYGNKLLGGY
           AAYRYAQPTATAAAYSDSYGRVYAADPYHHALAPATPYGVGAMNAPFLTDKTRSH
           ADDGVLSLSQASIYRGGRNRFAPY"
BASE COUNT      502 a 505 c 487 g 506 t
ORIGIN
Query Match      5.2%; Score 70; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1045 GATGCCAAGACTAGGAGCCATGCTGATGATGGTCTCGTCTCTTCTTCATTGCAGGCT 1104
|||||
Db 1411 GATGCCAAGACTAGGAGCCATGCTGATGATGGTCTCGTCTCTTCTTCATTGCAGGCT 1470
|||||

Qy 1105 AGTATATACC 1114
|||||
Db 1471 AGTATATACC 1480
|||||

RESULT 10
AF107203
LOCUS      2279 bp mRNA linear PRI 23-JUN-2000
DEFINITION Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.
ACCESSION  AF107203
VERSION     AF107203.1 GI:8671585
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 2279)
AUTHORS    Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE      A novel protein, A2BP, with RNA binding motif binds to C-terminal
           ataxin-2
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 2279)
AUTHORS    Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE      Direct Submission
JOURNAL    Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
           Laboratory, Burns and Allen Research Institute, Cedars-Sinai
           Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
           Angeles, CA 90048, USA
           Location/Qualifiers
FEATURES   1..2279
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /note="similar to Homo sapiens cDNA clone HSP38A20"
           1..2279
gene

```

```

CDS        /gene="A2BP"
           1047..2180
           /gene="A2BP"
           /codon_start=1
           /product="ataxin 2-binding protein"
           /protein_id="AAF78291.1"
           /db_xref="GI:8671586"
           /translation="MAQPYASAOFPAPNGIPAEVTAHPHPAPEYTGOTTVPHEITLN
           LYPPAOTHSSEQSPADTSAQTVSGTATQDDAAPTQGPOTQSPSTENKSKOPKRLHVS
           NIPFRDPLDROMFGQFGLDVEIIFNERSKGFVTFENSADADRAREKLGHTV
           VEGRKIEVNATARVMTNPKTVNTNGKLNPNVGVAVSPFYAGTIVLLCOAQEGS
           SMYSAPSSLYITISAMEGFYPAAATAAAYRGAAHLRGRGRTVNTFRAAAPPPPIPAYG
           GVYQDGFYADLYGGAAYRYAQPTATAAAYSDSYGRVYAADPYHHALAPATPYGVG
           GAMNAPFLTDKTRSHADDVGLVLSLSQASIYRGGRNRFAPY"
           1341..1358
           /misc_feature
           /gene="A2BP"
           /note="Region: RNA binding motif RNP-2"
           1452..1472
           /misc_feature
           /gene="A2BP"
           /note="Region: RNA binding motif RNP-1"
           497 a 712 c 636 g 434 t
BASE COUNT      497 a 712 c 636 g 434 t
ORIGIN
Query Match      5.2%; Score 70; DB 9; Length 2279;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1045 GATGCCAAGACTAGGAGCCATGCTGATGATGGTCTCGTCTCTTCTTCATTGCAGGCT 1104
|||||
Db 2079 GATGCCAAGACTAGGAGCCATGCTGATGATGGTCTCGTCTCTTCTTCATTGCAGGCT 2138
|||||

Qy 1105 AGTATATACC 1114
|||||
Db 2139 AGTATATACC 2148
|||||

RESULT 11
AR134676
LOCUS      2372 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6194171.
ACCESSION  AR134676
VERSION     AR134676.1 GI:14123581
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 2372)
AUTHORS    Pulst,S.M. and Shibata,H.
TITLE      Nucleic acids encoding ataxin-2 binding proteins
JOURNAL    Patent: US 6194171-A 1 27-FEB-2001;
           Location/Qualifiers
FEATURES   1..2372
           /organism="unknown"
           548 a 726 c 646 g 452 t
BASE COUNT      548 a 726 c 646 g 452 t
ORIGIN
Query Match      5.2%; Score 70; DB 6; Length 2372;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1045 GATGCCAAGACTAGGAGCCATGCTGATGATGGTCTCGTCTCTTCTTCATTGCAGGCT 1104
|||||
Db 2133 GATGCCAAGACTAGGAGCCATGCTGATGATGGTCTCGTCTCTTCTTCATTGCAGGCT 2192
|||||

Qy 1105 AGTATATACC 1114
|||||
Db 2193 AGTATATACC 2202
|||||

RESULT 12
HSFOX14
LOCUS      2971 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 alpha, beta, and

```

Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CAGCTGTGTG)
R. Site2: DraIII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method
(Sugano et al., Institute of Medical Science, University of Tokyo).

Custom primer used for sequencing
(5' end primer [CTTCTGCTCTAAAGCTGCG];
3' end primer [CGACCTGCAGCTCGACACA]).

FEATURES

source

Location/Qualifiers
1..1475

/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QtrA-11594"
/sex="male"
/tissue type="temporal lobe right"
/clone_lib="macaque brain cDNA library QtrA"
/dev_stage="adult"
90..1220

CDS

/codon_start=1
/product="hypothetical protein"
/protein_id="BAB46877.1"
/db_xref="GI:13874511"
/translation="MEERKGRVQGNQNEAAAPDTMAQPVASAFAPQNGIPAEYTP
APHPPAEYTGTTVPHTLNLYPPAQTHSEQSPADTNAQVSTGATQTDAAPTDG
QPTQPSNTENKSPQKRLHVSNIPIFRFPDLRMFGQFGKILDEVEIIFNERGSKGF
GFPTENSADADAREKLHGTVEGRKIEVNNATARVMTNKTVNPYTNWKLNPVVG
AVSPFEYAGTVLLCOAQEGSSMYSPSSLYVTSMPGFPVPAATAAAVYRGALRG
RGRTVYNTFRAAAPPFPPIPAYGGVYQDGFVGDYIGGYAARYAQPTATAAAYSDR
NQFVGADEISCNSTSAVTDFEMLPTPTITLLQPPPTALVP"

BASE COUNT 389 a 414 c 363 g 309 t

ORIGIN

Query Match 5.2%; Score 70; DB 9; Length 1475;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTTTCTTCATTGCAGGCT 1104

Db 1241 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTTTCTTCATTGCAGGCT 1300

QY 1105 AGTATATACC 1114

Db 1301 AGTATATACC 1310

RESULT 8

AK001027 1513 bp mRNA linear PRI 01-AUG-2002
Homo sapiens cDNA FLJ10165 fib, clone HEMBA1003591, weakly similar
to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.

AK001027

Accession AF229057.1 GI:7022045

Version AF229057

Keywords AF229057.1

Source AF229057.1

Organism Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,

Chen, W. and Winkelmann, J.C.

REFERENCE 1 (bases 1 to 2000)

AUTHORS Chen, W. and Winkelmann, J.C.

Nishikawa, T., Negai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Masuno, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5' - & 3' - end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

source

Location/Qualifiers
1..1513

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003591"
/tissue type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/note="cloning vector: pME18SFL3"
221..1414

CDS

/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91472.1"
/db_xref="GI:7022046"
/translation="MNCERQLRQNEAAAPDTMAQPVASAFAPQNGIPAEYTP
HPHPAEYTGTTVPHTLNLYPPAQTHSEQSPADTNAQVSTGATQTDAAPTDGQ
PTQPSNTENKSPQKRLHVSNIPIFRFPDLRMFGQFGKILDEVEIIFNERGSKGF
GFPTENSADADAREKLHGTVEGRKIEVNNATARVMTNKTVNPYTNWKLNPVVG
AVSPFEYAGTVLLCOAQEGSSMYSPSSLYVTSMPGFPVPAATAAAVYRGALRG
RGRTVYNTFRAAAPPFPPIPAYGGVYQDGFVGDYIGGYAARYAQPTATAAAYSDY
RYAADPVHHALAPPTYGVGMNAPLTDKTRSHADDVGLVLSLQASLIYRGYIN
RFAPY"

BASE COUNT 354 a 418 c 395 g 346 t

ORIGIN

Query Match 5.2%; Score 70; DB 9; Length 1513;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTTTCTTCATTGCAGGCT 1104

Db 1313 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTTTCTTCATTGCAGGCT 1372

QY 1105 AGTATATACC 1114

Db 1373 AGTATATACC 1382

RESULT 9

AF229057 2000 bp mRNA linear PRI 01-FEB-2002
Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
(HRNBP1) mRNA, complete cds.

AF229057

Accession AF229057.1 GI:18461366

Version AF229057

Keywords AF229057.1

Source AF229057.1

Organism Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2000)

AUTHORS Chen, W. and Winkelmann, J.C.

Consensus quality: 23573 bases at least Q40
 Consensus quality: 25350 bases at least Q30
 Consensus quality: 26823 bases at least Q20

```

* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/Senbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
      consists of 44 contigs. The true order of the pieces
      is not known and their order in this sequence record is
      arbitrary. Gaps between the contigs are represented as
      runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
      as soon as it is available and the accession number will
      be preserved.

```

*	1	1065:	contig of 1065 bp in length
*	1	1165:	gap of unknown length
*	1166	2411:	contig of 1246 bp in length
*	2412	2511:	gap of unknown length
*	2512	3982:	contig of 1471 bp in length
*	3983	4082:	gap of unknown length
*	4083	5596:	contig of 1514 bp in length
*	5597	5696:	gap of unknown length
*	5697	7081:	contig of 1385 bp in length
*	7082	7181:	gap of unknown length
*	7182	8496:	contig of 1315 bp in length
*	8497	8596:	gap of unknown length
*	8597	10100:	contig of 1504 bp in length
*	10101	10200:	gap of unknown length
*	10201	11249:	contig of 1049 bp in length
*	11250	11349:	gap of unknown length
*	11350	12701:	contig of 1352 bp in length
*	12702	12801:	gap of unknown length
*	12802	14212:	contig of 1411 bp in length
*	14213	14312:	gap of unknown length
*	14313	15380:	contig of 1068 bp in length
*	15381	15480:	gap of unknown length
*	15481	16518:	contig of 1036 bp in length
*	16517	16616:	gap of unknown length
*	16617	17946:	contig of 1330 bp in length
*	17947	18046:	gap of unknown length
*	18047	19146:	contig of 1100 bp in length
*	19147	19246:	gap of unknown length
*	19247	20226:	contig of 1280 bp in length
*	20227	20626:	gap of unknown length
*	20627	21830:	contig of 1104 bp in length
*	21831	21930:	gap of unknown length
*	21931	23358:	contig of 1528 bp in length
*	23359	23459:	gap of unknown length
*	23459	25338:	contig of 1880 bp in length
*	25339	25438:	gap of unknown length
*	25439	26638:	contig of 1200 bp in length
*	26639	26738:	gap of unknown length
*	26739	29368:	contig of 1630 bp in length
*	28369	28468:	gap of unknown length
*	28469	29896:	contig of 1428 bp in length
*	29897	29996:	gap of unknown length
*	29997	31050:	contig of 1054 bp in length
*	31051	31150:	gap of unknown length
*	31151	32585:	contig of 1435 bp in length
*	32586	32685:	gap of unknown length
*	32686	34092:	contig of 1407 bp in length
*	34093	34192:	gap of unknown length
*	34193	35491:	contig of 1299 bp in length
*	35492	35591:	gap of unknown length
*	35592	37096:	contig of 1505 bp in length
*	37097	37196:	gap of unknown length
*	37197	38312:	contig of 1116 bp in length
*	38313	38412:	gap of unknown length
*	38413	40579:	contig of 2167 bp in length
*	40580	40679:	gap of unknown length
*	40680	42711:	contig of 2032 bp in length
*	42712	42811:	gap of unknown length
*	42812	44224:	contig of 1413 bp in length

*	44225	44324:	gap of unknown length
*	44255	45859:	contig of 1535 bp in length
*	44325	45959:	gap of unknown length
*	45860	47975:	contig of 2016 bp in length
*	45960	48075:	gap of unknown length
*	47376	48049:	contig of 1974 bp in length
*	48076	50149:	gap of unknown length
*	50050	51700:	contig of 1551 bp in length
*	50150	51800:	gap of unknown length
*	51701	53551:	contig of 1751 bp in length
*	51801	53651:	gap of unknown length
*	53552	56127:	contig of 2476 bp in length
*	53652	56227:	gap of unknown length
*	56128	57920:	contig of 1693 bp in length
*	56228	58020:	gap of unknown length
*	57321	60257:	contig of 2237 bp in length
*	58021	60357:	gap of unknown length
*	60358	62838:	contig of 2481 bp in length
*	62839	62938:	gap of unknown length
*	62939	64839:	contig of 1901 bp in length
*	64840	64939:	gap of unknown length
*	64940	66877:	contig of 1938 bp in length
*	66878	66977:	gap of unknown length
*	66978	69856:	contig of 2879 bp in length
*	69857	69956:	gap of unknown length
*	69957	71716:	contig of 1760 bp in length
*	71717	71816:	gap of unknown length
*	71817	71929:	contig of 2113 bp in length

```

FEATURES
source
Location/Qualifiers
1. .73929
/organism="Rattus n
/db xref="taxon:101

```

BASE COUNT	ORIGIN	19038 a	16649 c	14558 g	19363 t	4321 others

Query Match 5.9%; Score 79; DB 2; Length 73929;
Best Local Similarity 100.0%; Pred. No. 4.8e-36;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 CAGTTACGGACGAGTTTATGTCTGCCGACCCCTACCACCAACACTTGCTCGAGCCCCCAC 1004
|||||
|||
Dk 11500 GACATTACCGGACGAGTTTATGTCTGCCGACCCCTACCACCAACACTTGCTCGAGCCCCCAC 11649
|||||

AB060859	AB060859	1475 bp	mRNA	linear	PRI 13-JUN-2000
LOCUS	Macaca fascicularis brain cDNA clone:QtrA-11594, full insert				
DEFINITION	sequence.				
ACCESSION	AB060859				
VERSION	AB060859.1	GI:13874510			
KEYWORDS	oligo capping; fis (full insert sequence).				
SOURCE	Macaca fascicularis adult male temporal lobe right cDNA to mRNa, clone lib.macamne brain cDNA library QtrA clone:QtrA-11594.				

ORGANISM
Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Macaca.

REFERENCE
AUTHORS
1 (sites)
Celcoprinaceinae; Macaca.
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao K., Suzuki Y., Sugano, S. and Hashimoto, K.

TITLE
Isolation of full-length cDNA clones from macaque brain cDNA libraries

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1475)
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan

```
QY 1263 GACATCACTTTAGCACTCAAAAGAAACAAC 1292
|||||
Db 112315 GACATCACTTTAGCACTCAAAAGAAACAAC 112286

RESULT 5
AF191501
LOCUS AF191501 1586 bp mRNA linear ROD 01-MAR-2002
DEFINITION Mus musculus hexaribonucleotide binding protein 1 (Hrnbp1) mRNA,
complete cds.
ACCESSION AF191501
VERSION AF191501.1 GI:19032413
KEYWORDS
SOURCE
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1586)
Chen, W. and Winkelman, J.C.
Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
Cincinnati
Location/Qualifiers
FEATURES
source
1..1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
gene 1..1586
/gene="Hrnbp1"
CDS 166..1419
/gene="Hrnbp1"
/codon_start=1
/product="hexaribonucleotide binding protein 1"
/db_xref="GI:19032414"
/translation="MLASQGVLLHSYGVPMIVPAAPYFPLGMLQNGQEAAAAAPDTMAQP
YASAFAPDDQAPAEYTAHPHPAPEYTGTTVPDHTLNLYPTQTTHSEQSADTSAQ
TVSGATQTDAAPTDGPQTPSPENTSKQPKLHVSNIPFRFRDPLQRMFOFG
KILDEIIFNFGSGKGFVTFENSADADRAKELHGVNIVSGKIEVNNATARYWTK
KTVNPTNGKLNPNVGVAYSPDFYAGTVLVCQANQEGSSMYSPSSLYTISMPGFP
YPAASAAAYRGALURGRGTGYNTLRAAAPPPPIPAYGVVGYDEPVTGNKLLQGGYA
AYRYAQPTPTAAAYSDSYGRVYAADPYHTLAPATPTYGVGMNAFAPLTDKTRSHA
DDVGLVLSLQASIVRGYNNRFAPY"
BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN

Query Match 12.0%; Score 161; DB 10; Length 1586;
Best Local Similarity 100.0%; Pred. No. 3.9e-86;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 ACCCTTACCAATGGCTGGAATTAATCCAGTGTGGCGCGGTCTACAGCCCGAC 615
|||||
Db 829 ACCCTTACCAATGGCTGGAATTAATCCAGTGTGGCGCGGTCTACAGCCCGAC 888
|||||

QY 616 TTCTATGACGACGCGTGTGTGTCAGGCAACACGAGGAGTCTTCCATGTACAGT 675
|||||
Db 889 TTCTATGACGACGCGTGTGTGTCAGGCAACACGAGGAGTCTTCCATGTACAGT 948
|||||

QY 676 GGGCCAGTTCATTGTATATATCTTCTGCAATGCCCTGGCTT 716
|||||
Db 949 GGGCCAGTTCATTGTATATATCTTCTGCAATGCCCTGGCTT 989
|||||

RESULT 6
AC103207
LOCUS AC103207 73929 bp DNA linear HTG 13-JUL-2002
DEFINITION Rattus norvegicus clone CH230-33J14, *** SEQUENCING IN PROGRESS
***, 44 unordered pieces.
ACCESSION AC103207
VERSION AC103207.3 GI:21731168
KEYWORDS HTG; HTGS_PHASE1.
```

SOURCE
ORGANISM

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 73929)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Albrechts, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Blom, D.,
Bouck, J., Bowie, S., Briefa, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssoen, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisege, H.,
Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.B., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoohtazi, N., Sisson, I.,
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umanak, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission

TITLE
JOURNAL

Unpublished
2 (bases 1 to 73929)
Worley, K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

3 (bases 1 to 73929)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17974697.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJLU
Center clone name: CH230-33J14
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990129

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Lucier, R., Luna, R., Ma, J.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwodu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sison, I.,
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, K., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Ugmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 115860)
 Worley, K.C.

Direct Submission
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 115860)
 Worley, K.C.

Direct Submission
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Jul 18, 2002 this sequence version replaced gi:20514522.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GXLG
 Center clone name: CH230-34C2
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 99693 bases at least Q40
 Consensus quality: 101936 bases at least Q30
 Consensus quality: 102743 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length.
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1347: contig of 1347 bp in length
 * 1348 1447: gap of unknown length
 * 1448 2834: contig of 1387 bp in length
 * 2835 2934: gap of unknown length
 * 2935 4253: contig of 1319 bp in length
 * 4254 4353: gap of unknown length
 * 4354 5477: contig of 1124 bp in length
 * 5478 5577: gap of unknown length
 * 5578 7584: contig of 1907 bp in length
 * 7585 9406: contig of 1822 bp in length
 * 9407 9506: gap of unknown length
 * 9507 11712: contig of 2206 bp in length
 * 11713 11812: gap of unknown length
 * 11813 13462: contig of 1650 bp in length
 * 13463 13562: gap of unknown length
 * 13563 17222: contig of 3660 bp in length
 * 17223 17322: gap of unknown length
 * 17323 20617: contig of 3295 bp in length
 * 20618 20717: gap of unknown length
 * 20718 24825: contig of 4108 bp in length
 * 24826 24925: gap of unknown length
 * 24926 28911: contig of 3886 bp in length
 * 28912 28911: gap of unknown length
 * 28912 33686: contig of 4775 bp in length
 * 33687 33786: gap of unknown length
 * 33787 40217: contig of 6431 bp in length
 * 40218 40317: gap of unknown length
 * 40318 47912: contig of 7595 bp in length
 * 47913 48012: gap of unknown length
 * 48013 56482: contig of 8470 bp in length
 * 56483 56582: gap of unknown length
 * 56583 66204: contig of 9622 bp in length
 * 66205 66304: gap of unknown length
 * 66305 74546: contig of 8242 bp in length
 * 74547 74547: gap of unknown length
 * 74548 86522: contig of 11876 bp in length
 * 86523 86622: gap of unknown length
 * 86623 100049: contig of 13427 bp in length
 * 100050 100149: gap of unknown length
 * 100150 115860: contig of 15711 bp in length.

FEATURES
 Location/Qualifiers
 source 1..115860
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-34C2"

BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others

ORIGIN

Query Match 12.5%; Score 168; DB 2; Length 115860;
 Best Local Similarity 99.3%; Pred. No. 2.5e-90;
 Matches 268; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1023 GAATGCTTTTGGCCCTTACCGGATGCCAGACTAGGAGCCATGCTGATGATGTTGGGTCT 1082
 Db 112555 GAATGCTTTTGGCCCTTACCGGATGCCAGACTAGGAGCCATGCTGATGATGTTGGGTCT 112496

Qy 1083 CGTCTTTTCTTCATTCAGGCTAGTATATACCAAGGGGATACAAACCGTTTGTCTCCATA 1142
 Db 112495 CGTCTTTTCTTCATTCAGGCTAGTATATACCGAGGGGATACAAACCGTTTGTCTCCATA 112436

Qy 1143 TTAATGATAAACCATTAAACAACAAGCAAAAAACAACAAAAACAAAAACCAAC 1202
 Db 112435 TTAATGATAAACCATTAAACAACAACAAGCAAAAAACAACAAAAACAAAAACCAAC 112376

Qy 1203 CTTCCAAATGTCGGAGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGACACATGAGTAG 1262
 Db 112375 CTTCCAAATGTCGGAGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGACACATGAGTAG 112316

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 14:22:28 ; Search time 3582.04 Seconds
(without alignments)
10887.019 Million cell updates/sec

Title: US-09-809-545A-1
Perfect score: 1340
Sequence: 1 gggcgccgcctgacacaaat.....ggccgaagggttcgctaga 1340

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sta.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_ot.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sta.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_mus.*
- 34: em_hgt_pln.*
- 35: em_hgt_rod.*
- 36: em_hgt_mam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_hgt_hum.*
- 40: em_hgt_mus.*
- 41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1340	100.0	1340	6	AX268800	AX268800 Sequence
2	194	14.5	1363	10	AF107204	AF107204 Mus muscu
3	194	14.5	2002	10	AB041596	AB041596 Mus muscu
4	168	12.5	115860	2	AC120661	AC120661 Rattus no
5	161	12.0	1586	10	AF191501	AF191501 Mus muscu
6	79	5.9	73929	2	AC103207	AC103207 Rattus no
7	70	5.2	1475	9	AB060859	AB060859 Macaca fa
8	70	5.2	1513	9	AK001027	AK001027 Homo sapi
9	70	5.2	2000	9	AF229057	AF229057 Homo sapi
10	70	5.2	2279	9	AF107203	AF107203 Homo sapi
11	70	5.2	2372	6	ARI34676	ARI34676 Sequence
12	70	5.2	2971	9	HSFOX14	AF109120 Homo sapi
13	70	5.2	3348	9	AF109106	AF109106 Homo sapi
14	70	5.2	164538	9	AC005774	AC005774 Homo sapi
15	67	5.0	115860	2	AC120661	AC120661 Rattus no
16	54	4.0	1547	9	AF094849	AF094849 Homo sapi
17	43	3.2	374	9	HSFOX08	AF109114 Homo sapi
18	43	3.2	1623	9	HSFOX08	AF109114 Homo sapi
19	40	3.0	234	9	HSFOX13	AF109119 Homo sapi
20	34	2.5	219	9	HSFOX05	AF109111 Homo sapi
21	34	2.5	112976	9	AC027683	AC027683 Homo sapi
22	34	2.5	169765	9	AC006075	AC006075 Homo sapi
23	32	2.4	629	9	HSFOX03	AF109109 Homo sapi
24	32	2.4	173824	2	AC095278	AC095278 Rattus no
25	31	2.3	169089	9	AC008537	AC008537 Homo sapi
26	31	2.3	169192	2	AC114672	AC114672 Mus muscu
27	31	2.3	217280	2	AC019337	AC019337 Homo sapi
28	31	2.3	224930	2	AC121911	AC121911 Mus muscu
29	31	2.3	235183	2	AC079427	AC079427 Mus muscu
30	30	2.2	5695	6	AX251292	AX251292 Sequence
31	30	2.2	5898	6	AX278016	AX278016 Sequence
32	30	2.2	5898	6	AX323715	AX323715 Sequence
33	30	2.2	5898	6	AX346842	AX346842 Sequence
34	30	2.2	6126	6	AX251328	AX251328 Sequence
35	30	2.2	6126	6	AX346733	AX346733 Sequence
36	30	2.2	6154	6	AX346207	AX346207 Sequence
37	30	2.2	6510	6	AX251401	AX251401 Sequence
38	30	2.2	6510	6	AX251810	AX251810 Sequence
39	30	2.2	6510	6	AX345296	AX345296 Sequence
40	30	2.2	57759	9	ALI37162	ALI37162 Human DNA
41	30	2.2	134787	2	AL356692	AL356692 Homo sapi
42	30	2.2	203279	9	AL354826	AL354826 Human DNA
43	30	2.2	260534	2	AC078912	AC078912 Mus muscu
44	29	2.2	219	9	HSFOX04	AF109110 Homo sapi
45	29	2.2	1076	9	BC020939	BC020939 Homo sapi

ALIGNMENTS

RESULT 1	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
LOCUS	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
DEFINITION	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
ACCESSION	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
VERSION	AX268800.1	GI:16541860				
KEYWORDS	Norway rat.					
SOURCE	Rattus norvegicus					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;					
	Rattus.					
REFERENCE	Stanton,L.W. and White,R.T.					
AUTHORS	Secreted factors					
TITLE						

[illegible]

```

RESULT 14
US-09-789-561-55
; Sequence 55, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 55
; LENGTH: 1198
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-789-561-55

```

	Query Match	3.7%	Score 49.2	DB 10	Length 1198
	Best Local Similarity	54.4%	Pred. No. 0.011		
	Matches 99	Conservative 0	Mismatches 83	Indels 0	Gaps 0
Qy	1137	TCCTATTAAATGATAAAACCATTTAAACAAACAAGCAAAAAACAAAAACAAAAACAAAAA	1196		
Db	1011	TCMAATATAACTGGTAGCAACGTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAA	1070		
Qy	1197	ACCAACCTTCCAAATGTGGGGAGAGAGGAGCTTCCCGAGGCCCGAGTGTGGGACACATG	1256		
Db	1071	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAA	1130		
Qy	1257	CAGTAGGACATCACCTTTAGCAACTCAAAGAAACAACGAAAAAATAAAAAAATAAAAAA	1316		
Db	1131	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAA	1190		
Qy	1317	AA 1318			
Db	1191	AA 1192			

RESULT 15
US-09-983-965-2109/c
; Sequence 2109, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26

```

, PRIOR APPLICATION NUMBER: US 09/465,231
,
, PRIOR FILING DATE: 1999-12-15
, PRIOR APPLICATION NUMBER: US 60/113,678
, PRIOR FILING DATE: 1998-12-17
, NUMBER OF SEQ ID NOS: 5912
, SEQ ID NO 2109
, LENGTH: 529
, TYPE: DNA
, ORGANISM: Bos taurus
, FEATURE:
, NAME/KEY: unsure
, LOCATION: (335)
, OTHER INFORMATION:
, OTHER INFORMATION: Clone ID: 34-LIB3057-015-Q1-K1-A6
US-09-983-965-2109

```

	Query Match	3.6%	Score 48.8	DB 10	Length 529
	Best Local Similarity	54.9%	Pred. No. 0.0091		
	Matches	95	Conservative	0	Mismatches 78
					Indels 0
					Gaps 0
Qy	1146	ANTGATAAACCATTTAAACAAACAAAGCAAAACAAACAAACAAACAAACAAACAAACAAACCTT	1205		
Db	461	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	402		
Qy	1206	CCAATGTGGGAGAGAGGAAGCTTTCCGAGGCCGAGTGTTCGCACACATGCAGTAGGAC	1265		
Db	401	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	342		
Qy	1266	ATCACTTTAGCACTCAAGAAACAACGAAAAAATAAATAAATAAATAAATAAATAAATAAATAA	1318		
Db	341	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	289		

Search completed: March 15, 2003, 16:03:57
Job time : 132.943 secs

```
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4237 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-745-763-20

Query Match          3.9%; Score 52; DB 10; Length 4237;
Best Local Similarity 55.8%; Pred. No. 0.0038;
Matches 100; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1150 ATAAACCATTAACAAACGCAAAACAAACAAACAAACAAACAAACCAACCTTCCAA 1209
Db 2430 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2489

QY 1210 TGTGGGAGAGAGAGCTTCCGAGCGCGAGTGTTCGACACATGCATGAGCATCA 1269
Db 2490 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2549

QY 1270 CTTTAGCACTCAAGAAACAAACGAAACAAACAAACAAACAAACAAACAAACG 1329
Db 2550 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2609

RESULT 11
US-10-001-843-45
; Sequence 45, Application US/10001843
; Patent No. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; CURRENT APPLICATION NUMBER: US/10/001,843
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 1267
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (358)..(358)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc feature
; LOCATION: (478)..(478)
; OTHER INFORMATION: a, c, g or t
US-10-001-843-45

Query Match          3.7%; Score 49.8; DB 12; Length 1267;
Best Local Similarity 55.5%; Pred. No. 0.0077;
Matches 96; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1146 AATGATAAACCATTAACAAACGCAAAACAAACAAACAAACAAACAAACCAACCTT 1205
Db 976 AAAAAAAAAAAAAAAAAACAAACGAAAGAAAAAAAAAAAAAAAAACAAACAAACAA 1035
```

```
QY 1206 CCAATGTGGGAGAGAGAGCTTCCGAGGCCGCGAGTGTTCGACACATGCATGAGGAC 1265
Db 1036 CAAAGAAAGAGAGAGAGGAAAAATAAGAGCAAAAAACAAAAAAGAAAAATGAC 1095

QY 1266 ATCACTTTAGCACTCAAGAGAAACACGAAAAAAGAAAAAAGAAAAAAGAAAAAATA 1318
Db 1096 AAAAAACAGAAAAAACAAGATACAAACAAACAAAGAAAGAAACAAACAAAGAA 1148

RESULT 12
US-09-960-352-573/c
; Sequence 573, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 573
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 03-LIB3058-031-Q1-K1-A11
US-09-960-352-573

Query Match          3.7%; Score 49.4; DB 10; Length 428;
Best Local Similarity 54.7%; Pred. No. 0.0058;
Matches 98; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1140 ATATTAAATGATAAACCATTTAAACAAACAGCAACAAACAAACAAACAAACAAAC 1199
Db 407 ATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATA 348

QY 1200 AACCTTCCATGTGGGAGAGAGAGCTTCCGAGGCCGCGAGTGTTCGACACATGCAG 1259
Db 347 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAATATAAAAAAAAAAGAAATAA 288

QY 1260 TAGGACATCACTTTAGCACTCAAGAGAAACAAACGAAAAAAGAAAAAAGAAAAA 1318
Db 287 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAAAAAAAA 229

RESULT 13
US-09-960-352-3400/c
; Sequence 3400, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3400
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 15-LIB3058-052-Q1-K1-D11
US-09-960-352-3400

Query Match          3.7%; Score 49.4; DB 10; Length 446;
Best Local Similarity 54.7%; Pred. No. 0.006;
Matches 98; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

Db 289 TNGTNNNTNTTATNTTTCTTTCTTTTNTGAAATTAATAAAGGNAANAAAAA 230

Qy 1131 TTTTGTCTCCATATTAAATGATAAACCATTAACAAACAAAGCAAAAAACAAAAAC 1190

Db 229 ANAANTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 170

Qy 1191 AAAAAAACCACTTCCAATGTGGGAGAGAGAGCTTTCCGAGGCCGAGTTTCGGA 1250

Db 169 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 110

Qy 1251 CACATGCGAGTAGGACATCACTTTAGCACTCAAGAGACACGAAAAAATAAATAA 1310

Db 109 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 50

Qy 1311 AAAAAATAA 1318

Db 49 AAAAAATAA 42

RESULT 8

US-09-825-294-53/c

Query Match 3.9%; Score 52.8; DB 10; Length 396;

Best Local Similarity 48.4%; Pred. No. 0.00079;

Matches 120; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun

APPLICANT: Stolk, John A.

APPLICANT: Algate, Paul A.

APPLICANT: Fling, Steven P.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER

FILE REFERENCE: 210121.484CS

CURRENT APPLICATION NUMBER: US/09/825.294

CURRENT FILING DATE: 2001-04-03

NUMBER OF SEQ ID NOS: 215

SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 53

LENGTH: 396

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(396)

OTHER INFORMATION: n = A,T,C or G

US-09-825-294-53

Qy 1071 TCATGTGGGTCTCGTCTCTTTCTTTCATTGCGGCTAGTATATACCAAGGGGGATACACCG 1130

Db 289 TNGTNNNTNTTATNTTTCTTTCTTTTNTGAAATTAATAAAGGNAANAAAAA 230

Qy 1131 TTTTGTCTCCATATTAAATGATAAACCATTAACAAACAAAGCAAAAAACAAAAAC 1190

Db 229 ANAANTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 170

Qy 1191 AAAAAAACCACTTCCAATGTGGGAGAGAGAGCTTTCCGAGGCCGAGTTTCGGA 1250

Db 169 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 110

Qy 1251 CACATGCGAGTAGGACATCACTTTAGCACTCAAGAGAAACAAAGAAAAAATAAATAA 1310

Db 109 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 50

Qy 1311 AAAAAATAA 1318

Db 49 AAAAAATAA 42

RESULT 9

US-09-960-352-4582/c

Query Match 3.9%; Score 52; DB 10; Length 393;

Best Local Similarity 53.4%; Pred. No. 0.0013;

Matches 109; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengbing

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960.352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 4582

LENGTH: 393

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 20-LIB3057-013-Q1-K1-E7

US-09-960-352-4582

Qy 1122 ATACAACCGTTTGTCTCATATTAAATGATAAAACCACTTAAACAAACAAAGCAAAAAACAA 1181

Db 213 ATAAAAAATTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 154

Qy 1182 AACAAAAACAAAAAACCAACCTTCCAATGTGGGAGAGAGAGCTTTCCGAGGCCGA 1241

Db 153 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 94

Qy 1242 GTGTGGCAGACATGCGAGTAGGACATCACTTTAGCACTCAAGAGAAACAAAGAAAAA 1301

Db 93 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 34

Qy 1302 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1325

Db 33 AAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 10

RESULT 10

US-09-745-763-20

Query Match 3.9%; Score 52; DB 10; Length 393;

Best Local Similarity 53.4%; Pred. No. 0.0013;

Matches 109; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.

APPLICANT: LaVallie, Edward R.

APPLICANT: Collins-Racie, Lisa A.

APPLICANT: Evans, Cheryl

APPLICANT: Marberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745.763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 28375
;; LENGTH: 125
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC006075.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.67
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.7
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.5
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
;; OTHER INFORMATION: EST HUMAN HIT: 225303.1, EVALUATE 9.00e-61
;; OTHER INFORMATION: NT HIT: g11431054, EVALUATE 2.00e-64
;; OTHER INFORMATION: SWISSPROT HIT: P54537, EVALUATE 1.40e+00
;; US-09-864-761-28375

Query Match 8.2%; Score 109.6; DB 10; Length 125;
Best Local Similarity 92.7%; Pred. No. 2.8e-18;
Matches 115; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 239 CGACGAGCCGAGCCGACAGACACCTTCTGAAACACAGAAACAGTCCAGCCCA 298
Db 1 CGACGGATGGCCGAGCCGACAGACACCTTCTGAAACACGGAACAGTCTCAGCCCA 60
Qy 299 AGCGGCTGATGTCTCAACATCCCTTCGCGGTTCGGGATCCGACCTCCGACAAATGT 358
Db 61 AGCGGCTGATGTCTCCATATATCCCTTCAGGTTCCGGATCCGACCTCAGACAAATGT 120
Qy 359 TTGG 362
Db 121 TTGG 124

RESULT 5
US-09-864-761-22653

;; Sequence 22653, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzei, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;; FILE REFERENCE: Aeomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 22653
;; LENGTH: 224
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC006075.1
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
;; OTHER INFORMATION: NT HIT: g18922072, EVALUATE 1.00e-123
;; OTHER INFORMATION: SWISSPROT HIT: Q09457, EVALUATE 8.00e-03
;; OTHER INFORMATION: EST_HUMAN HIT: AU118435.1, EVALUATE 1.00e-123
;; US-09-864-761-22653

Query Match 6.6%; Score 88; DB 10; Length 224;
Best Local Similarity 82.6%; Pred. No. 9.4e-13;
Matches 185; Conservative 0; Mismatches 30; Indels 9; Gaps 7;

QY 652 CAGGAGGATCTTCCATGTACAGTGGCCCGCAGTTCACTTGTATATATCTTCTGCAATGCCT 711
| | | | |
Db 1689 CAGGAGGATCTTCCATGTACAGTGGCCCGCAGTTCACTTGTATATATCTTCTGCAATGCCT 1748
| | | | |
QY 712 GGCTTTCCATATCGGGCGGCACCTGTGAGCTGTCATACGAGGGGTCACTTCGAGGC 771
| | | | |
Db 1749 GGCTTTCCATATCGGGCGGCACCTGTGAGCTGTCATACGAGGGGTCACTTCGAGGC 1808
| | | | |
QY 772 CGTGGTCGACCGTGTAAACACCTTCAGAGCTGCGGCGCCCGCCACCCCAATCCCGGCC 831
| | | | |
Db 1809 CGGGTCGACCGTGTAAACACCTTCAGGGCGCGGCGCCCGCCCGCCGATCCCGGCC 1868
| | | | |
QY 832 TATGGCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATAAATGGCTACAGGGTGGTTAC 891
| | | | |
Db 1869 TACGGCGGTGTGTTTACC--CAGGTGATTTTATGGTCGACACATTTATGTTGGTTAT 1926
| | | | |
QY 892 GCTGCATACCGCTACGCCCGACCCCTGCGACCTGCTGCTGCTGCTACAGTGA----- 944
| | | | |
Db 1927 GCTGCATACCGCTACGCCCGACCCCTACCCCTGCCACTGCGCTGCTACAGTGACAAAT 1986
| | | | |
QY 945 -----CAGTTACGGACGAG 958
| | | | |
Db 1987 CAGTTCGTCTTCTGTGAGCAGATGAATTTCTTTAAACCTCTCGAGTTACGGACGAG 2046
| | | | |
QY 959 TTTATGTCGCGACCCCTTACCAACACACACTTGTCTCAGCGCCCGCCACCTACGGGCTTGGTG 1018
| | | | |
Db 2047 TTTATGTCGCGACCCCTTACCAACACACACTTGTCTCAGCGCCCGCCACCTACGGGCTTGGTG 2106
| | | | |
QY 1019 CCATGAATGCTTTTGGCCCTTGCACCGATGCCAAGACTAGGAGCCATGCTGTATGATGTGG 1078
| | | | |
Db 2107 CCATGAATGCTTTTGGCCCTTGCACCGATGCCAAGACTAGGAGCCATGCTGTATGATGTGG 2166
| | | | |
QY 1079 GTCTCGTCTTTCTTCAATTCAGCTAGTATATACCAAGGGGGATCAACCGTTTGCCTC 1138
| | | | |
Db 2167 GTCTCGTCTTTCTTCAATTCAGCTAGTATATACCAAGGGGGATCAACCGTTTGCCTC 2226
| | | | |
QY 1139 CATATTAAATGATAAAACCACTTAAACAAACAAAGCAAAACAAACAAACAAACAAAC 1198
| | | | |
Db 2227 CATACTAAATGACAAACCACTAAA----- 2250
| | | | |
QY 1199 CAACCTTCCAAATGTGGGAGAGAGAGAGCTTTCGAGGCGCGAGTGTGGACACATGCA 1258
| | | | |
Db 2251 -AACCTTCCAAATGTGGGAGAGAGAGCTTTCGAGGCGCGAGTGTGGACACATGCA 2309
| | | | |
QY 1259 GTAGGACATCACTTTAGCAACTCAACGACACACGAAACAAACAAACAAACAAACAAAT 1318
| | | | |
Db 2310 GTAGTACATCACTTTAGCAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 2369
| | | | |

RESULT 3
US-09-864-761-11804
; Sequence 11804, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11804
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006075.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.67
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
US-09-864-761-11804

Query Match 9.2%; Score 123.8; DB 10; Length 473;
Best Local Similarity 91.6%; Pred. No. 1.4e-21;
Matches 131; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 220 CAGACAGATGATCCCGCCGACCGACCGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 279
| | | | |
Db 269 CAGACAGATGATCCCGCCGACCGACCGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 328
| | | | |
QY 280 GAAACAAAGTCCCGCCGACCGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 339
| | | | |
Db 329 GAAACAAAGTCCCGCCGACCGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 388
| | | | |
QY 340 CCAGACCTCCGACAAATGTTTG 362
| | | | |
Db 389 CCGGACCTCAGACAAATGTTTG 411
| | | | |

RESULT 4
US-09-864-761-28375
; Sequence 28375, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23

Qy 301 CGGCTGCAATGTGTCAACATCCCTTCCGTTCCGGATCCAGACCTCCGACAAATGTTT 360
Db 301 CGGCTGCAATGTGTCAACATCCCTTCCGTTCCGGATCCAGACCTCCGACAAATGTTT 360
Qy 361 GGCCTATTTGTAATAATATTAGATGTTGAAATATTATTTAATAGCGGGCTCGAAGGA 420
Db 361 GGCCTATTTGTAATAATATTAGATGTTGAAATATTATTTAATAGCGGGCTCGAAGGA 420
Qy 421 TTTGGTTTCGTAACCTTCGAAAAATAGTCGGATCGGACAGCGGCGAGGAGAAATTGCAC 480
Db 421 TTTGGTTTCGTAACCTTCGAAAAATAGTCGGATCGGACAGCGGCGAGGAGAAATTGCAC 480
Qy 481 GGTACCGGTGAGAGGCGGTAATAATCGAGGTTAATAATGCGACAGCAGCGGTGATGACT 540
Db 481 GGTACCGGTGAGAGGCGGTAATAATCGAGGTTAATAATGCGACAGCAGCGGTGATGACT 540
Qy 541 AATAAAGAGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTGGGCGG 600
Db 541 AATAAAGAGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTGGGCGG 600
Qy 601 GTCTACAGCCCGACTTCTATGAGCGACGCTGCTGTTGTGCGAGGCCAAACAGAGGGA 660
Db 601 GTCTACAGCCCGACTTCTATGAGCGACGCTGCTGTTGTGCGAGGCCAAACAGAGGGA 660
Qy 661 TCTTCATGTACAGTGGCCCGAGTTCACTTGTATATATCTTCTGCAATGCTGGCTTCCA 720
Db 661 TCTTCATGTACAGTGGCCCGAGTTCACTTGTATATATCTTCTGCAATGCTGGCTTCCA 720
Qy 721 TATCGGGCGGCACTGTGAGCTGCATACCGAGGGGCTCACCTTCGAGGGCGGTGCTGC 780
Db 721 TATCGGGCGGCACTGTGAGCTGCATACCGAGGGGCTCACCTTCGAGGGCGGTGCTGC 780
Qy 781 ACCGTGTAACACACTTCAGAGCTCGCGCGCCCAACCCCAATCCCGGCTATGGCGGA 840
Db 781 ACCGTGTAACACACTTCAGAGCTCGCGCGCCCAACCCCAATCCCGGCTATGGCGGA 840
Qy 841 GTAGTGTATCAAGAGCCAGTGTATGGCAATAAATGTCTACAGGGTGTACGCTGCATAC 900
Db 841 GTAGTGTATCAAGAGCCAGTGTATGGCAATAAATGTCTACAGGGTGTACGCTGCATAC 900
Qy 901 CGTACGCCCGGACCCCACTGCTGCTGCTACAGTGTACAGTGTACGAGGAGTT 960
Db 901 CGTACGCCCGGACCCCACTGCTGCTGCTACAGTGTACAGTGTACGAGGAGTT 960
Qy 961 TATGCTGCGGACCCCTACCCACACACTTGTCTCAGGCCCCCACCCTACGGGCTTGGTGC 1020
Db 961 TATGCTGCGGACCCCTACCCACACACTTGTCTCAGGCCCCCACCCTACGGGCTTGGTGC 1020
Qy 1021 ATGAATGCTTTTTCGCCCTTGAACCGATGCAAGACTAGGAGCCATGCTGATGTGGGT 1080
Db 1021 ATGAATGCTTTTTCGCCCTTGAACCGATGCAAGACTAGGAGCCATGCTGATGTGGGT 1080
Qy 1081 CTCGTTCTTTCTTCAATGAGGCTAGTATATACCAAGGGGGATACAACTGTTTCTCTCA 1140
Db 1081 CTCGTTCTTTCTTCAATGAGGCTAGTATATACCAAGGGGGATACAACTGTTTCTCTCA 1140
Qy 1141 TATTAAATGATAAACCATTAAACAAACAGCAAAACAAACAAACAAACAAACAAACAA 1200
Db 1141 TATTAAATGATAAACCATTAAACAAACAGCAAAACAAACAAACAAACAAACAAACAA 1200
Qy 1201 ACCTTCCAATGTGGGAGAGAGAGCTTTCGAGGCGCGAGTGTGTGGACACATGCACT 1260
Db 1201 ACCTTCCAATGTGGGAGAGAGAGCTTTCGAGGCGCGAGTGTGTGGACACATGCACT 1260
Qy 1261 AGGACATCACTTTAGCACTCAAGAAACAAACGAAAAAATAAGC 1320
Db 1261 AGGACATCACTTTAGCACTCAAGAAACAAACGAAAAAATAAGC 1320
Qy 1321 GGCAGAGGGGTTCCCTAGA 1340
Db 1321 GGCAGAGGGGTTCCCTAGA 1340

RESULT 2
US-09-794-591-1
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CR 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987)..(1979)
US-09-794-591-1
Query Match 64.0%; Score 857.2; DB 10; Length 2372;
Best Local Similarity 82.7%; Pred. No. 7.5e-205;
Matches 1141; Conservative 0; Mismatches 138; Indels 101; Gaps 10;
Qy 1 GCGGCGGCGCTGACAAATGGCTCAG-CTTATGCTTCAGGCGAGTTGCTTCACCCCGAG 59
Db 1029 GCGGCTGCGCTTGACAAATGGCTCAGCTTACGCTTCGGCCAGTTTCTCCCGCGAG 1088
Qy 60 AATGGCATCT-GCAGAATACAGCGCCCTCATCCCATCCCGCCGAGAG-ACACCGGC 117
Db 1089 AACGGTATCCCGCGGAATACAGCGCCCTCATCCCATCCCGCCGAGAGTACACAGC 1148
Qy 118 CAGCCCACTGTCCCG-CCACACATTAACCTTG-ATCCTCTACACAGACGACCTCGGAG 175
Db 1149 CAGACACGGTTCCCGAGCAGACATTAACCTGTACCTCCCGCCGAGACGACCTCGGAG 1208
Qy 176 CAGAGC--GCTTATACAGGCGACAG-CGCTCTCCGCGACCGCCACACACAGACATGAT 231
Db 1209 CAGAGCGCGCGGACAGAGCGCTCAGACCGCTCTCTGSCACCGCCACACACAGACATGAC 1268
Qy 232 GCGCCCGCGAGCGCGCCAGCCCGACACAACTTCTGAAACACAGAAACAAAGTCC 291
Db 1269 GCAGCAGCGAGTGGCCAGCCCGCCAGACACAACTTCTGAAACACAGAAACAAAGTCT 1328
Qy 292 CAGCCCAAGCGCTGCATGTGTCAACATCCCTTCCGGTTCCGGGATCCAGACCTCCGA 351
Db 1329 CAGCCCAAGCGCTGCATGTGTCTCAATATCCCTTCAGTTCCGGGATCCGAGCTCAGA 1388
Qy 352 CAATGTTTGGCCAAATTTGGTAAATATATAGATGTTGAAATATATTTTAAATAGCGGGGC 411
Db 1389 CAATGTTTGGTCAATTTGGTAAATATCTAGATGTTGAAATATATTTTAAATAGCGGGGC 1448
Qy 412 TCGAGGGATTTGCTTTCGTAACTTTCGAAATAGTCGGATCGGACAGCGCGGGAG 471
Db 1449 TCAAGGGATTTGCTTTCGTAACTTTCGAAATAGTCGGATCGGACAGCGGGAGGAG 1508
Qy 472 AAATTGACCGGTACCGGTGAGAGGCGGTAAATTCAGAGTTAAATTCGACAGCAGCGC 531
Db 1509 AAATTACGCGTACCGGTGAGAGGCGGTAAATTCAGAGTTAAATTCGACAGCAGCGT 1568
Qy 532 GTGATGACTAATAAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTT 591
Db 1569 GTAATGACAAATAAAGACCGTCAACCCCTTATACAAATGCTGGAATTAATTCAGTT 1628
Qy 592 GTGGCGGGCTTACAGCCCGGACTTCTATCAGGACGCGGTGCTGTTGTCAGCGCCAC 651
Db 1629 GTGGGTGAGTCTACAGTCCCGAATTTCTATCAGCAGCGGCTCTGTTGTGCGAGGCCAAC 1688

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:44:58 ; Search time 118.943 Seconds
(without alignments)
7907.139 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340

Sequence: 1 gggcgccgcctgacacaat.....ggccgaagggttcgctaga 1340

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 501302 seqs, 350932545 residues

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1340	100.0	1340	10	US-09-809-545A-1
2	857.2	64.0	2372	10	US-09-794-591-1
3	123.8	9.2	473	10	US-09-864-761-11804
4	109.6	8.2	125	10	US-09-864-761-28375
5	88	6.6	224	10	US-09-864-761-22653
6	86.4	6.4	454	10	US-09-864-761-5893
7	52.8	3.9	396	10	US-09-970-966-53
8	52.8	3.9	396	10	US-09-825-294-53
9	52	3.9	4237	10	US-09-745-763-20
10	52	3.9	4237	10	US-10-001-843-45
11	49.8	3.7	1267	12	US-09-960-352-573
12	49.4	3.7	428	10	US-09-960-352-3400
13	49.4	3.7	1198	10	US-09-789-561-55
14	49.2	3.7	1198	10	US-09-983-965-2109
15	48.8	3.6	529	10	US-09-960-352-2223
16	48.4	3.6	415	10	US-09-960-352-2223
17	47.8	3.6	200	10	US-09-960-352-4816
18	47.8	3.6	312	10	US-09-960-352-8414
19	47.8	3.6	425	10	US-09-960-352-4010

C 20	47.6	3.6	2305	10	US-09-795-006A-124	Sequence 124, App
C 21	47.4	3.5	217	10	US-09-960-352-6539	Sequence 6539, App
C 22	47.4	3.5	283	10	US-09-960-352-9095	Sequence 9095, App
C 23	47.2	3.5	259	10	US-09-960-352-7233	Sequence 7233, App
C 24	47.2	3.5	463	10	US-09-960-352-7186	Sequence 7186, App
C 25	47	3.5	237	10	US-09-960-352-14543	Sequence 14543, A
C 26	46.8	3.5	241	10	US-09-960-352-12302	Sequence 12302, A
C 27	46.6	3.5	242	10	US-09-960-352-3873	Sequence 3873, App
C 28	46.2	3.4	207	10	US-09-960-352-3946	Sequence 3946, App
C 29	46.2	3.4	344	10	US-09-960-352-1036	Sequence 1036, App
C 30	46.2	3.4	439	10	US-09-960-352-1009	Sequence 1009, App
C 31	46.2	3.4	491	10	US-09-878-574-4301	Sequence 4301, App
C 32	46.2	3.4	516	10	US-09-960-352-5785	Sequence 5785, App
C 33	46.2	3.4	4167	10	US-09-764-878-282	Sequence 282, App
C 34	46.2	3.4	4167	10	US-09-764-860-1145	Sequence 1145, App
C 35	46.2	3.4	4167	10	US-09-764-846-345	Sequence 345, App
C 36	46	3.4	202	10	US-09-764-846-280	Sequence 280, App
C 37	46	3.4	936	10	US-09-739-907-47	Sequence 47, Appl
C 38	45.6	3.4	390	10	US-09-960-352-1976	Sequence 1976, App
C 39	45.6	3.4	831	10	US-09-729-674-67	Sequence 67, Appl
C 40	45.4	3.4	430	10	US-09-960-352-9747	Sequence 9747, App
C 41	45.2	3.4	545	10	US-09-878-574-4299	Sequence 4299, App
C 42	45.2	3.4	741	10	US-09-910-943-432	Sequence 432, App
C 43	45.2	3.4	758	10	US-09-910-943-490	Sequence 490, App
C 44	45.2	3.4	760	10	US-09-910-943-366	Sequence 366, App
C 45	45	3.4	448	10	US-09-878-574-4308	Sequence 4308, App

ALIGNMENTS

RESULT 1

US-09-809-545A-1
; Sequence 1, Application US/09809545A
; Patent No. US20020110804A1
; GENERAL INFORMATION:
; APPLICANT: Stanton, Lawrence W.
; APPLICANT: White, R. Tyler
; TITLE OF INVENTION: SECRETED FACTORS
; FILE REFERENCE: SCIOS.017A
; CURRENT APPLICATION NUMBER: US/09/809,545A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1340
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-809-545A-1

Query Match 100.0%; Score 1340; DB 10; Length 1340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CGCGCGCCCTGACACAAATGGCTCAGCTTATCGCTCAGGCGAGTTCGCTCCACCCGAGA	60
DB	1	CGCGCGCCCTGACACAAATGGCTCAGCTTATCGCTCAGGCGAGTTCGCTCCACCCGAGA	60
QY	61	ATGGCATCTTCGAGATACACGGCCCTCATCCCATCCCGCCGACAGACACGGCCGAG	120
DB	61	ATGGCATCTTCGAGATACACGGCCCTCATCCCATCCCGCCGACAGACACGGCCGAG	120
QY	121	CCCACTGTCTCCCGCCACACATTAACTTGATCTCTCTACACAGACGCACTCGGAGCAGAG	180
DB	121	CCCACTGTCTCCCGCCACACATTAACTTGATCTCTCTACACAGACGCACTCGGAGCAGAG	180
QY	181	CGCTTATACAGCGCAAGCGGTCTCGGCAACCGCCACACAGACAGATGATCGGCCCGG	240
DB	181	CGCTTATACAGCGCAAGCGGTCTCGGCAACCGCCACACAGACAGATGATGCGGCCCGG	240
QY	241	ACGAGCGGCGCCCGCCACAGACAACTTCTGAAAAACACAGAAAAAAGTCCCGAGCCCAAG	300
DB	241	ACGAGCGGCGCCCGCCACAGACAACTTCTGAAAAACACAGAAAAAAGTCCCGAGCCCAAG	300


```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3

Query Match      3.2%; Score 42.4; DB 3; Length 7286;
Best Local Similarity 58.9%; Pred. No. 0.068;
Matches 73; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 1139 CATATTAAATGATAAAACCAATTAAACAAAGCAAGAAACAAACAAACAAACAAACAAACAAAC 1198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4613 CAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4672

Qy 1199 CAACCTTCCAATGTGGGAGAGAGAGAGCTTCCGAGGCCCGAGTGTTCGACACATGCA 1258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4673 AAAAAAGTACCTTCTGAGCGGGAAGAACCCAGCCGGATCCAGACATGATAAGATACATTGA 4732

Qy 1259 GTAG 1262
    ||
Db 4733 TGAG 4736

RESULT 15
US-09-331-581-14
; Sequence 14, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-0PCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 14
; LENGTH: 7938
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14

Query Match      3.2%; Score 42.4; DB 3; Length 7938;
Best Local Similarity 58.9%; Pred. No. 0.072;
Matches 73; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 1139 CATATTAAATGATAAAACCAATTAAACAAAGCAAGAAACAAACAAACAAACAAACAAAC 1198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5265 CAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5324

Qy 1199 CAACCTTCCAATGTGGGAGAGAGAGCTTCCGAGGCCCGAGTGTTCGACACATGCA 1258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5325 AAAAAAGTACCTTCTGAGCGGGAAGAACCCAGCCGGATCCAGACATGATAAGATACATTGA 5384

Qy 1259 GTAG 1262
    ||
Db 5385 TGAG 5388
```

Search completed: March 15, 2003, 16:01:35
Job time : 134.316 secs

QY 335 GGGATCAGACCTCCGCAAAATGTTGGCCAAATTTGGTAAATAATTAGATGTTGAAATTA 394
DB 377 SNGARGAYGAYTNAARACNGTNTTYGCNCARTTYGGNGCNGTNGTNGARGTNAAYATHC 436
QY 395 TTTTAA--TGAGCGGGCTCGAAGGATTTGGTTTCGTAACCTTTCGAAAATAGTCGG 451
DB 437 CNMGNAARCCNGAYGNAARATGMNGGNTTYGNTTYGNCARTTYGAARAAAYTYNTNG 496
QY 452 ATCGGACAGCGGCGAGGAGAAATTCACCGTACCGTGTAGAGGCGCGTAAATCGAGG 511
DB 497 ARGCGNGAARGCNYTNAARGGNATGAAYATGAARGARATHAARGGNMGNCNGTNGCNG 556
QY 512 TTAATAATCGACAGCAGCGGTGATGACTAATAAAA 547
DB 557 TNGAYTGGCGNGTNGCNAARGAAYAAATAYAAARGAYA 592

RESULT 12
US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-098-487-2

Query Match 3.4%; Score 45; DB 2; Length 2277;
Best Local Similarity 32.4%; Pred. No. 0.0064;
Matches 109; Conservative 46; Mismatches 178; Indels 3; Gaps 1;
QY 215 CCACACAGACAGATGATGCGCCCGCAGCGGCGGAGCCCGCCAGACACCTTCTGAAA 274
DB 257 CNAARGAARGGNAARAAAYGARAAYNSGARTGYCCNAARARGARCCNAARGCNAARA 316
QY 275 ACACAGAAAAAAGTCCCGACCCAGCGGTGTCATGTGTCCAAACATCCCTTCGGGTTC 334
DB 317 ARGCNAARGTNGCNGAYAAARAARCCNMGNTNATHATMGNAAYTYNTTVAARTGYW 376
QY 335 GGGATCAGACCTCCGCAAAATGTTGGCCAAATTTGGTAAATAATTAGATGTTGAAATTA 394
DB 377 SNGARGAYGAYTNAARACNGTNTTYGCNCARTTYGGNGCNGTNGTNGARGTNAAYATHC 436

QY 395 TTTTAA--TGAGCGGGCTCGAAGGATTTGGTTTCGTAACCTTTCGAAAATAGTCGG 451
DB 437 CNMGNAARCCNGAYGNAARATGMNGGNTTYGNTTYGNCARTTYGAARAAAYTYNTNG 496
QY 452 ATCGGACAGCGGCGAGGAGAAATTCACCGTACCGTGTAGAGGCGCGTAAATCGAGG 511
DB 497 ARGCGNGAARGCNYTNAARGGNATGAAYATGAARGARATHAARGGNMGNCNGTNGCNG 556
QY 512 TTAATAATCGACAGCAGCGGTGATGACTAATAAAA 547
DB 557 TNGAYTGGCGNGTNGCNAARGAAYAAATAYAAARGAYA 592

RESULT 13
US-09-342-653-5
; Sequence 5, Application US/09342653
; Patent No. 6306632
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Chromatin Associated Proteins
; FILE REFERENCE: BB-1118
; CURRENT APPLICATION NUMBER: US/09/342,653
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: 60/092,841
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-342-653-5

Query Match 3.2%; Score 43; DB 4; Length 572;
Best Local Similarity 55.8%; Pred. No. 0.011;
Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
QY 1088 TTTCTTCATTGCGAGGTAGTATATACCAAGGGGTATACCAACCGTTTGTCTCCATATTTAA 1147
DB 426 TTTTCTTCATCCCTTCTATTGCTGAAGAAGATACCGTTGGGAAAAAATAAAAA 485
QY 1148 TGATAAAACCATTAACAAACAAAGCAAAAAACAAAAACAAAAACCAACCTTCC 1207
DB 486 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 545
QY 1208 ATGTGGGGAGAGAGAGCTTTCCGA 1234
DB 546 AAAATCTCGAGGGGGCGCGGTACCCA 572

RESULT 14
US-09-331-581-3
; Sequence 3, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-OPCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7286

Query Match 3.4%; Score 45.2; DB 4; Length 2394;
Best Local Similarity 47.8%; Pred. No. 0.0058;
Matches 131; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 3 GGCGCCCTGACACATGGCTGCTATGCTTATGCTCAGCGCAGTTCGCTCCACCCAGAAAT 62
DB 1742 GGTGCTCCCATCCACACCTTACACCCACCCCGGAGCTGATGATGTGAGCTGTG 1801

QY 63 GGCAATCTGACAGATACACGGCCCTCATCCCCATCCCGCCAGAGACACCGGCCAGCC 122
DB 1802 GGCGGCTGCTGCTGCTCCCACTTACACCGGCTGCTCCCGCCAGCAGCCGCTGCT 1861

QY 123 CACTGTCCCGCCACACATTAATCTGATCTCTACACAGCAGCTCGGAGCAGAGCG 182
DB 1862 CAGCCCTCCGAGCTCGATGACTGAGGCTGCTCCACCCCTCCCGCTCCTGATGACCC 1921

QY 183 CTTATACAGCGCAGCGCTCGGACCGCCACACAGCAGATGATGCGGCCGCGAC 242
DB 1922 CCACTCTGGGGCTTACCTGCGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1981

QY 243 CGAGCGCCAGCCCGCCAGACACAACTTCTGAAAC 276
DB 1982 ACCCTGCCCCCTCCCTGGGGCCCTCTGAAAGC 2015

RESULT 10
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-2

Query Match 3.4%; Score 45; DB 1; Length 2277;
Best Local Similarity 32.4%; Pred. No. 0.0064;
Matches 109; Conservative 46; Mismatches 170; Indels 3; Gaps 1;

QY 215 CCACAGACAGATGATGCGCCCGCCAGCGCGCCAGCCCGCCAGACACAACTTCTGAA 274
DB 257 CNAARGAARAGGNAARAAAYGARAAYWSNGARTGYCCNNAARAGARCCNNAARAG 316

QY 275 ACACAGAAACAAGTCCAGCCCAAGCGGCTGATGTGTCAACATCCCTTCGGTTCC 334
DB 317 ARGCAARGTNGCNGAYAAARAGCNGMNYTNATHATMGNAAYYTNWSNTTYAARTGYW 376

QY 335 GGGATCCAGACCTCCGACAAATGTTTGGCCAAATTTGGTAAATAATTAGATGTTGAAATTA 394
DB 377 SNGARGAYGYNNAARACNGTNTTYGCNARTTYGGNGCNGTNYNGARGTNAAYATHC 436

QY 395 TTTTAA---TGAGCGGGGCTCGAAGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGG 451
DB 437 CNMGNAARCCNGAYGNAARATGMNGGNTTYGNTTYGTNCARTTYAARAAAYTNTYNG 496

QY 452 ATGGGACAGCGGAGGAGAAATTCACCGGTACCGTGGTAGAGGGCCGTAATAATCGAGG 511
DB 497 ARGCGNNAARCGNYNNAARAGGNATGAAYATGAARGARATHAARGGNGMNGACNGTNGCG 556

QY 512 TTAATAATGCGACAGCAGCGCTGATGACTAATAAAA 547
DB 557 TNGAYTGGGCTGNGCNAARAGAYAAARTAYAAARGAYA 592

RESULT 11
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-2

Query Match 3.4%; Score 45; DB 1; Length 2277;
Best Local Similarity 32.4%; Pred. No. 0.0064;
Matches 109; Conservative 46; Mismatches 178; Indels 3; Gaps 1;

QY 215 CCACAGACAGATGATGCGCCCGCCAGCGCGCCAGCCCGCCAGACACAACTTCTGAA 274
DB 257 CNAARGAARAGGNAARAAAYGARAAYWSNGARTGYCCNNAARAGARCCNNAARAG 316

QY 275 ACACAGAAACAAGTCCAGCCCAAGCGGCTGATGTGTCAACATCCCTTCGGTTCC 334
DB 317 ARGCAARGTNGCNGAYAAARAGCNGMNYTNATHATMGNAAYYTNWSNTTYAARTGYW 376

db 926 ACCCTGCCCCCTCCCTGGGGCCCCCTCCTGAAAGC 959

RESULT 7

```

US-09-271-815-3
; Sequence 3, Application US/09271815
; Patent No. 6297036
; GENERAL INFORMATION:
; APPLICANT: BERGSMÄ, DERK
; APPLICANT: SHABON, USMAN
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
; FILE REFERENCE: GH-70172-1
; CURRENT APPLICATION NUMBER: US/09/271,815
; CURRENT FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 09/027,064
; EARLIER FILING DATE: 1998-02-20
; EARLIER APPLICATION NUMBER: 60/053,924
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-271-815-3

```

Query Match	3.4%	Score 45.2;	DB 4;	Length 1338;
Best Local Similarity	47.8%;	Pred. No. 0.0041;		
Matches 131;	Conservative 0;	Mismatches 143;	Indels 0;	Gaps 0;

QY	3	GGCGCGCCCTGACACAATGGCTCAGCTTATGGCTCAGCGAGTTGCTTCCACCCCAAGAT	62
Db	686	GGTCGTCCCCATCCAAACCTCACCACCCCGGAGCTGATGGATGTGAGCCTGGTG	745
QY	63	GGCATCTGTGAGNAATACACGGCCCTCATCCCCATCCGCGCCGAGACACCGGCCAGCC	122
Db	746	GGCGGCCTGTGATGCTGTCCCACTCACCAGCGCCTGCCCCCAGCACCCGGCTGGCT	805
QY	123	CACGTGTCGCCGCCACACATTAAACTTGATCTCTTACACAGACGCACTCGGAGCAGAGCG	182
Db	806	CAGCCCTCCGACCTGGNATGACTGGAGTGCTCACCCCTCTCCGCGCTCTGATGACCTG	865
QY	183	CTTATACAAGCGCACAGCCGCTTCGGGCACCGCCACACACAGATGATGCGGCCCGAC	242
Db	866	CCACTCTGGGSCCTCACCTGGSCCTCCGTGGTGTAACCCAGAGACACAGCAGCCAGTCGTG	925
QY	243	CGACGGCGAGCCCGACAGACAACTTTGAAAAAC	276
Db	926	ACCTTGCCCCCTCTCTGGGGGCCCTCTCTGAAGC	959

RESULT 8

```

US-09-027-064-1
; Sequence 1, Application US/09027064
; Patent No. 6133006
; GENERAL INFORMATION:
; APPLICANT: SHABON, USMAN
; APPLICANT: BERGSMAN, DEREK
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE
; TITLE OF INVENTION: PROTEIN KINASE-HTLAR33
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

```

; SOFTWARE: FastSeq for Windows Version 2.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,064
; FILING DATE: 20-FEB-1998
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/053,924
; FILING DATE: 28-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2394 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-09-027-064-1

```

Query Match	3.4%	Score 45.2;	DB 3;	Length 2394;
Best Local Similarity	47.0%;	Pred. No. 0.0058;		
Matches 131;	Conservative	0;	Mismatches 143;	Indels 0; Gaps 0;
QY	3	GGCGGCCCTGACACATGGCTAGCTTATGCCTCAGCGCAGTTTCGCTCCACCCAGAAAT	62	
Db	1742	GGTGTGCTCCCATCACCACCCCTCACCACACCCCGGAGTGATGGATGTGAGCGCTGGTG	1801	
QY	63	GGATCTCTGAGAAATACAGGCCCTCATCCCATCCGCGCCAGAGACACCGGCCAGCC	122	
Db	1802	GGCGGCTGTGACTGTCTCCCACTCACCAGGCGTGCCTCCAGCACCAGCGTGCCT	1861	
QY	123	CAGTGTCCCCGGCACATTAACCTTCCTACACAGACGCAGCTCGGAGCAGAGCG	182	
Db	1862	CAGCCCTCCGACTCGGATGACTGGAGTGTCCACCCCTCCCGCTCCTGATGACCCTG	1921	
QY	183	CTTATACAAGGCACAGCCGCTCTCGGCACCGCCACACAGACAGATGATCGGCCCCGAC	242	
Db	1922	CCACTCTGGGCGCTCACTGGCCTCGTGGTGTACCCACAGACAGCAGCAGCTCGTG	1981	
QY	243	CGAGGGCAGCCCGACACAACTTCTGAAAAC	276	
Db	1982	ACCTGCCCCCTCCTGGGCGCCCTCTGAAAGC	2015	

RESULT 9

```

US-09-271-815-1
; Sequence 1, Application US/09271815
; Patent No. 6297036
; GENERAL INFORMATION:
; APPLICANT: BERGSMÄ, DERK
; APPLICANT: SHABON, USMAN
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
; FILE REFERENCE: GH-70172-1
; CURRENT APPLICATION NUMBER: US/09/271.815
; CURRENT FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 09/027,064
; EARLIER FILING DATE: 1998-02-20
; EARLIER APPLICATION NUMBER: 60/053,924
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2394
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-271-815-1

```



```
Qy 352 CAAATGTTTGCCCAATTTGGTAAATAATTAGATGTTGAAATTAATTTTAAATGACGGGC 411
Db 1389 CAAATGTTTGGTCAATTTGGTAAATCTTAGATGTTGAAATTAATTTTAAATGACGGGC 1448
Qy 412 TCGAAGGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGATGCGACAGGCGAGGAG 471
Db 1449 TCAAGGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGATGCGACAGGCGAGGAG 1508
Qy 472 AAATTTGACCGGTACCGTGTAGAGGCGGTAAATTCGAGGTTAATAATCGACAGCAGCG 531
Db 1509 AAATTTACCGGACCGTGTAGAGGCGGTAAATTCGAGGTTAATAATCGACAGCAGCG 1568
Qy 532 GTGATGACTAATAAAGAGCGGTGAACCCCTACACCAATGCTGGAATTAATAATCCAGTT 591
Db 1569 GTAATGACAAATAAAGACCGTCAACCCCTTATACAAATGCTGGAATTAATAATCCAGTT 1628
Qy 592 GTGGCGGGCTTACAGAGCCCGACTTCTATGAGGACAGGTCGTTGTGCGAGGCCAAC 651
Db 1629 GTGGGTGAGTCTACAGTCCCGAATTTCTATGAGGACAGGTCGTTGTGCGAGGCCAAC 1688
Qy 652 CAGGAGGATCTTCATGTACAGTGGCCCGCAGTTCACTTGTATATACCTCTGCAATGCCT 711
Db 1689 CAGGAGGATCTTCATGTACAGTGGCCCGCAGTTCACTTGTATATACCTCTGCAATGCCT 1748
Qy 712 GGCTTTCCATATCCGGCGCCCACTGCTGACGTGTCATACCGAGGGGCTCACCTTCGAGGC 771
Db 1749 GGCCTTCCGTATCCAGAGCCACCGCGCGCGCTACCGAGGGGCGCACCTCGCGAGGC 1808
Qy 772 CGTGTGCGACGGTGTACACACCTTACAGAGTGGGGCGCGCCCGCCCAATCCCGGCC 831
Db 1809 CGCGTGCACCGGTGTACACACCTTACAGGCGCGCGCGCGCCCGCCCGCTCCCGGCC 1868
Qy 832 TATGCGGAGTAGTGTATCAAGAGCCAGTGTATGCAATAAATTCCTACAGGTTGGTTAC 891
Db 1869 TAGCGCGTGTGTTTACC--CAGATGATTTTATGTTGACAGATTTATGTTGTTAT 1926
Qy 892 GCTGCATACCGCTACGCCCGAGCCCAACCCCTGCCACTGCTGCTGCCTACAGTGA----- 944
Db 1927 GCTGCATACCGCTACGCCCGAGCCTACCCCTGCGCTGCGCTGCTACAGTGACAGAAAT 1986
Qy 945 -----CAGTTACGACGAG 958
Db 1987 CAGTTGCTTTCGTTGACGACAGATGAAATTTCTTGTAAACACCTCTGCGAGTTACGGACGAG 2046
Qy 959 TTTATGCTGCCGACCCCTACACACACACACTTGTCTCCAGCCCGCCACCTACGCGTGTGTG 1018
Db 2047 TTTATGCTGCCGACCCCTACACACGACCTTGTCTCCAGCCCGCCACCTACGCGTGTGTG 2106
Qy 1019 CCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGGAGCCATGCTGATGTGG 1078
Db 2107 CCATGAATGCTTTTGCACCTTTGACTGATGTCGAAGCTAGGAGCCATGCTGATGTGG 2166
Qy 1079 GTCTCGTCTTTTCTTCATTCGAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTC 1138
Db 2167 GTCTCGTCTTTTCTTCATTCGAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTC 2226
Qy 1139 CATATTAATGATAAAACCATTTAAACAAACAGCAAAAAAACAACAAAAA 1198
Db 2227 CATACTAATGCAAAACCATTA----- 2250
Qy 1199 CAACCTTCCAATGTGGGAGAGAGGAGCTTTCCAGGCGCCGAGTGTTCGACACATGCA 1258
Db 2251 -AACCTTCCAATGTGGGAGAGAGGAGCTTTCCAGGCGCTGAGTATTGCAATACATGCA 2309
Qy 1259 GTAGACATCACTTTAGCAACTCAAGAAACAAAGAAAAAATAA 1318
Db 2310 GTAGTACATCACTTTAGCAATTTTAAAAAATAAATAAATAAATAAATAA 2369
```

RESULT 2

US-09-176-657-5

; Sequence 5, Application US/09176657

```
; Patent No. 6020164
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lu, Aina
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
; FILE REFERENCE: P2-0611 US
; CURRENT APPLICATION NUMBER: US/09/176,657
; CURRENT FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1250374
US-09-176-657-5
```

Query Match 20.6%; Score 275.6; DB 3; Length 1506;

Best Local Similarity 60.2%; Pred. No. 8.5e-69;

Matches 601; Conservative 0; Mismatches 354; Indels 43; Gaps 7;

```
Qy 222 GACAGATGATGCGCCCGCCGACCGACGCGCCAGC-----CCGACACACAACTTCTGAAAAAC 278
Db 452 GACAGAGGTGAGCACAGACAGACAGCGCCAGCAGTCAAGACACAAAGTAGTGAATTC 511
Qy 279 AGAAAAAAGTCCAGCGCCAAAGCGGTCATGTGTCAACATCCCTTCGGGTTCCGGGA 338
Db 512 AGAGAGTAAATCTACCCCGAAACGGCTGCTCTCTAATAATTCTTCCGCTTCCGGGA 571
Qy 339 TCCAGACTCCGACAAATGTTTGCCCAATTTGGTAAATATTAGATGTTGAAATTAATTTT 398
Db 572 CCCTGACTCCGGCAGATGTTTGGGCAGTTTGGCAGAAATCCTAGATGTAGAAATAATCTT 631
Qy 399 TAATGAGCGGGCTCGAAGGGAATTTGGTTTCGTAACCTTTCGAAAAATAGTGCAGTGGGA 458
Db 632 TAATGAACGTGGCTCTAAGGGATTCGGGTTTCGTAACCTTTCGAGAAATAGTGTGATGAGA 691
Qy 459 CAGGCGAGGAGGAGAAATGACGGTACCGGTAGCGGCGGTAAATTCGAGGTTTAATAA 518
Db 692 CAGGCGAGGAGGAGAAATATACAGCGCACCGTGTAGAGGGCGGTAAAAATCGAGGTGAATAA 751
Qy 519 TGCACAGCAGCGCTGATGACTAATAAAGCGCGTGAACCCCTACACCAATGGCTGGA 578
Db 752 TGCTACAGCAGCTGTAATGACCAATAAGAAATGGTCAACATATGCAATGGTTGGA 811
Qy 579 ATTAATCCAGTTTGTGGCGCGGCTTACAGCCCGGACTTCTATGCG-----GCA 628
Db 812 ATTAAGCCCACTAGTTGAGAGCTGATATGTTCCGGAGTTATATGACGATCCAGCTTCA 871
Qy 629 CGGTGCTGTTGTCAGCGCCAAACAGAGGGAATCTTCCATGTA--CAGTGGCCCCAGTTC 686
Db 872 AGCAGATGTTGTCCTAGGCAATGATGACAGCAGTGCCTTATCAGAGAGAGGGGATCAAA 931
Qy 687 ACTGTATATCTTCTGCAATGCTGGCTTTCCATA---TCGGCGCGCAGCTGCTGCGAGC 743
Db 932 CACTTACATTCCTTTAATCATCTTCTGGCTTCCCTTACCTACTGAGCCACACCGGAGC 991
Qy 744 TGCATACCGAGGGGCTCACCTTCGAGGCGCGTGGTGCACCGGTGACAAACACTTTCAGAGC 803
Db 992 CGCTTTCAGAGAGGCCCAATTTGAGGGGCGAGGGCGGACAGTATATGTTGAGTCCGA-- 1049
Qy 804 TGGGCGGCCCGCCACCCCAATCCCGGCTATGGCGGAGTAGTGTATCAAGAGCCAGTGTGA 863
Db 1050 -GCGGTACTCTCAACAGCCATCCCGCCTATCCAGGGGTGGATATGACAGCCTACAG---- 1104
Qy 864 TGGCAATAAATGCTACAGGGTGGTTAGCTGCATACCGCTACCGCCAGCCCGCCCTGCG 923
Db 1105 -----ATATGACAGCCTGCTACTGCAACCGGAGCCACCGCTGCTGCGAGC 1149
```


THIS PAGE BLANK (USPTO)

DEFINITION ge72h11.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
IMAGE:1744581 3' similar to TR:Q10572 Q10572 44.4 KD RNA-BINDING
PROTEIN IN FOX-1 LOCUS. ; mRNA sequence.

ACCESSION A1193843

VERSION A1193843.1 GI:3745052

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 451)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

AUTHORS Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

JOURNAL This clone is available royalty-free through LNL; contact the
COMMENT IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 849 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 370.

FEATURES
Location/Qualifiers
1..451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1744581"
/clone_lib="Soares fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W." 110 g 120 t 1 others

BASE COUNT 107 a 113 c 110 g 120 t 1 others
ORIGIN

Query Match 28.2%; Score 378.2; DB 9; Length 451;
Best Local Similarity 91.2%; Pred. No. 4.7e-58;
Matches 412; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY 295 CCCAAGCGGCTGCATGTGTCACACATCCCTCCGGTTCGGGATCCAGACCTCCGACAA 354
|||
Db 451 CCCAAGCGGCTGCATGTGTCACATATCCCC-TCAGGTTCCGGATCCGGACCTCAGACAA 393
|||

QY 355 ATGTTTGGCCAAATTTGGTAAATATTAGATGTTGAAATATTTTAAATAGCGGGGCTCG 414
|||
Db 392 ATGTTTGGTCAATTTGGTANAATCTTAGATGTTGAAATATTTTAAATAGCGGGGCTCA 333
|||

QY 415 AAGGGATTTGGTTTCGTAATCTTCGAAATAGTCGGATCCGACAGCGGCGAGGAGAA 474
|||
Db 332 AAGGGATTTGGTTTCGTAATCTTCGAAATAGTCGGATCCGACAGCGGCGAGGAGAA 273
|||

QY 475 TTGCACGGTACCGTGTAGAGGCGCCGTAAATCGAGGTTAAATGCGACAGCAGCGGTG 534
|||
Db 272 TTACACGGCACCGTGTAGAGGCGCCGTAAATCGAGGTTAAATGCGACAGCAGCGGTGTA 213
|||

QY 535 ATGACTAATAAAGCGCGTGAACCCCTACACCAATGGCTGGAATTAATCCAGTTGTG 594
|||
Db 212 ATGACAAATAAAGAGACCGTCAACCCCTATACAAATGGCTGGAATTAATCCAGTTGTG 153
|||

QY 595 GCGCGGTCTACAGCCCGCTTCTATGAGGCGACCGTGTGTCGCGGCGCAACCCAG 654
|||
Db 152 GGTGCAGTCTACAGTCCCGAATTTCTATGAGGCGACCGTGTGTCGCGGCGCAACCCAG 93
|||

QY 655 GAGGGATCTTCCATGTCAGTGGCCCCAGTTCACCTTGTATATATCTTCTGCAATGCCTGGC 714
|||
Db 92 GAGGGATCTTCCATGTCAGTGGCCCCAGTTCACCTTGTATATATCTTCTGCAATGCCTGGC 33
|||

QY 715 TTTCATATCCCGCGCCGACACTGCTGCAGCTGC 746
|||
Db 32 TTCCCGTATCCAGAGCACCCCGCGCGCGC 1
|||

Search completed: March 15, 2003, 15:58:44
Job time: 1897.53 secs

```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-C-04-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p7T73D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

```

BASE COUNT	119 a	156 c	100 g	85 t
ORIGIN				
Query Match		28.6%	Score 383.2;	DB 9; Length 460;
Best Local Similarity		98.0%;	Pred. No. 6e-59;	
Matches 451; Conservative		0; Mismatches 3;	Indels 6; Gaps 6;	

Qy	2	CGCGCGCCCTGACACAAATGGCTCAG - CTTATGCTCAGCGCAGTTGCTCCACCCGAGA	60
Db	1	CGCGCGCCCTGACACAAATGGCTCAGCGCTTATGCTCAGCGCAGTTGCTCCACCCGAGA	60
Qy	61	ATGSCAT - CTTGCAGAATACACGCGCCCTCATCCCATCCCGCGCCGACAG - ACACCGGCC	118
Db	61	ATGSCATCCCTGCGAATACACGCGCCCTCATCCCATCCCGCGCCGAGTACACCGGCC	120
Qy	119	AGCCCACTGTCCCG - CCACACATTTAAACTTGG - ATCTCTCTACACAGACGCACTCGGAGC	176
Db	121	AGACCACTGTCCCGACCCACACATTTAAACCTGTATCTCTTACACAGACGCACTCGGAGC	180
Qy	177	AGAGGGCTTATACAAGCGCAG - CCGTCTCCGGCACCGCCACACAGACAGATGATGCCG	235
Db	181	AGAGGGCTGATACAAGCGCAGACGCGTCTCCGGCACCGCCACACAGACAGTATGCGG	240
Qy	236	CCCGACCGCAGCGGCAGCCCGACACAACTTTCTGAAAAACACAGAAACCAAGTCCCGCAGC	295
Db	241	CCCGACCGCAGCGGCAGCCCGACACAACTTTCTGAAAAACACAGAAACCAAGTCCCGCAGC	300
Qy	296	CCAGCGGCTGCATGTGTCCAACATCCCGTTCGGGATCCGGGATCCAGACCTCCGACAAA	355
Db	301	CCAGCGGCTGCATGTGTCCAACATCCCGTTCGGGATCCGGGATCCAGACCTCCGACAAA	360
Qy	356	TGTTTGGCCAATTTGGTAAAAATTATAGATGTTGAAATTTATTTTAAATGAGCGGGGCTCGA	415
Db	361	TGTTTGGCCAATTTGGTAAAAATTATAGATGTTGAAATTTATTTTAAATGAGCGGGGCTCGA	420
Qy	416	AGGGATTTGGTTTCGTAACTTTTCGAAAATATAGTCGGGATGC	455
Db	421	AGGGATTTGGTTTCGTAACTTTTCGAAAATATAGTCGGGATGC	460

RESULT 14	AV6665025	Bos taurus	brain fetus	549 bp	mrna	linear	EST 28-NOV-2001
LOCUS	AV665025	Bos taurus	clone E1BR045H01				
DEFINITION	5', mRNA sequence.						
ACCESSION	AV665025						
VERSION	AV665025.1	GI:9924055					
KEYWORDS	EST.						
SOURCE	cow.						
ORGANISM	Bos taurus						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.						
AUTHORS	1 (bases 1 to 549) Takasuga,A., Hirotsune,S., Itoh,R., Jitchazon,A., Suzuki,H., Aso,H. and Sugimoto,Y.						
TITLE	Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs						

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
 MEDLINE Contact: Yoshikazu Sugimoto
 COMMENT Animal Genetics Division
 Shirakawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5641
 Fax: 81-248-25-5725
 Email: kazusugi@cocoa.ocn.ne.jp
 Single pass sequencing.
 This clone was obtained from a polyA-deleted cDNA library.
 FEATURES Location/Qualifiers
 source 1..549
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone="E1BR045H01"
 /clone_lib="Bos taurus brain fetus"
 /tissue_type="brain"
 /dev_stage="fetus"
 /lab_host="DH108"
 /notes="Vector: pZL1; Site_1: Sali; Site_2: NotI; Poly A
 was deleted from a NotI site"
 BASE COUNT 151 a 151 c 141 g 104 t 2 others
 ORIGIN

Query Match	28.3%	Score 379.8	DB 10	Length 549	
Best Local Similarity	84.3%	Pred. No. 2.3e-58			
Matches 463	Conservative	0	Mismatches 79	Indels 7	Gaps 3
QY	84	CCCTCTCCTCCATCCCGCGCAGAGACACGGGCGACGCCACCTGTCTCCCG-CCACACATT	142		
Db	1	CCACGGTCCGCCACGGTCCGGAGTACATGGCCACACACCGTCCCGGAGCACAGTT	60		
QY	143	AAACTTG-ATCTCTCTACACAGACGCACTCGGAGCAGAGCGCTTATACAGCGCACAG--	199		
Db	61	AAACTGTACTCTCCCGCGAGTCCGCACTCGGAGCAGAGCGCCGNGGACACTGAGCGCG	120		
QY	200	---CGGTCTCCGCGACCGCCACACAGACAGATGATCGCCCGACCGAGCGCCAGCCCC	256		
Db	121	ACACCGTCTCGGGCACGGCCACACAGACAGATGACTCGGCACCCCGATGCCAGCCCC	180		
QY	257	AGACACAACTTCTGAAAAACACAGAAAAACAAGTCCCGAGCCCAAGCGGTGCATGTCTCA	316		
Db	181	AGACACAGCTTCTGAAAAACACGGAACAAGTCCCGAGCCCAAGAGGTGCATGTCTCA	240		
QY	317	ACATCCCTTCCGGTTCGGGATCCGACCTCCGACAAAATGTTTGGCCAAATTCGTAAAA	376		
Db	241	ACATCCCTTCCGGTTCGGGATCCGACCTCCGACAAAATGTTTGGTCAATTTGGTAAAA	300		
QY	377	TATTAGATGTTGAAATTAATTTTAAATGAGCGGGCTCGAAGGGATTTGGTTTCGTAACCT	436		
Db	301	TCTTAGATGTTGAAATTAATTTTAAATGAGCGAGGCTCAAGGGATTTGGTTTCGTAACCT	360		
QY	437	TCGAAAATAGTTCGGATGCGGACAGGGCGAGGAGAAAATTCACGGTACCGTGGTAGAGG	496		
Db	361	TCGAAAATAGTTCGGATGCGGACAGGGCGAGGAGAAAATTCACGGCACCGTGGTAGAGG	420		
QY	497	GCGGTAAATC3AGGTTAATTAATGGGACAGCACCGGTGATGACTAATAAAGGCCGTGA	556		
Db	421	GCGGTAAATC3AGGTTAATTAATGGACAGCACGTTGTAATGACAAATAAAGAACTGTGA	480		
QY	557	ACCCCTACACCAATGGCTGGAAAATTAATTCAGTTGTGGCGCGGTCTFACAGCCCCGACT	616		
Db	481	ACCCCTTATACAAATGGCTGGAAAATTAATTCAGTTGTGGGTGCAGNGTACAGTCTCTGAAT	540		
QY	617	TCTATGCAG 625			
Db	541	TCTATGCAG 549			
RESULT 15					
AI193843/c					
LOCUS	AI193843	451 bp	mRNA	linear	EST 29-OCT-1998

RESULT 15
AI193843/
LOCUS

```
QY 1013 TTGGTGCATGAATGCTTTTGGCCCTTGACCATGCGCAAGACTAGAGCCATGCTGATG 1072
Db 298 TTGGTGCATGAATGCTTTTGGCCCTTGACCATGCGCAAGACTAGAGCCATGCTGATG 357
QY 1073 ATGTGGGTCTCGTCTTTCTTCAATGCGAGCTAGTATATACCAAGGGGATACAAACCGTT 1132
Db 358 ATGTGGGTCTCGTCTTTCTTCAATGCGAGCTAGTATATACCAAGGGGATACAAACCGTT 417
QY 1133 TTGCTCCATATTAAATGATATAAAACCAATTAACAAACGCAAAACAAACAAACAA 1192
Db 418 TTGCTCCATATTAAATGATATAAAACCAATTAACAAACGCAAAACAAACAAACAA 447
QY 1193 AAAAACCAACCTTCCAAATGTTGGGAGAGAGAAAGCTTTCCGAGCCCGAGTGTTCGACA 1252
Db 448 -----AACCTTCCAATGTGGGAGAGAAAGCTTTCCGAGCCCGAGTATTGCAATA 500
QY 1253 CATGCGTAGGACATCATCTTTAGCAACTCAAGNACAAACGCAAAACAA 1299
Db 501 CATGCGTAGTATCATCTTTTAGCAACTCTAAACAAACGCAAAACAAACAA 547

RESULT 12
AV729198 773 bp mRNA linear EST 17-OCT-2000
LOCUS AV729198 HTC Homo sapiens cDNA clone HTCBCD09 5', mRNA sequence.
ACCESSION AV729198
VERSION AV729198.1 GI:10838619
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
AUTHORS Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA HTC clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1. .773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBCD09"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 184 a 232 c 185 g 170 t 2 others
ORIGIN

Query Match 28.7%; Score 384.2; DB 10; Length 773;
Best Local Similarity 82.1%; Pred. No. 3.3e-59;
Matches 482; Conservative 0; Mismatches 65; Indels 40; Gaps 2;

QY 713 GCTTTTCATATCGCGCGGCACTGCTGCAGCTGCATACCGAGGGGTCACTTCGAGGCC 772
Db 1 GCTTCCGTTATCCAGCAGCACCGTCGCGCGCCCTACCGAGGGGCGCACCTTCGAGGCC 60
QY 773 GTGGTCGCGCCGTGTACACACCTTTCAGAGCTCGCGGGCCCCCACCCCAATCCCGGCT 832
Db 61 GCGGTGCGCCCGTGTACAAACCTTTCAGGCGCGCGGGCCCCCGCCCGGCT 120
```

```
QY 833 ATGGCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATAAAATTGCTACAGGTTGGTTACG 892
Db 121 ACGGCGGTGTGTTTACCAGGATGGATTTTATGG---TGCAGACATTTATGGTGGTTATG 177
QY 893 CTGCATACCGCTACGCCAGCCAGCCACCCCTGCCACTGCTGCTACAGTGACAGTTACG 952
Db 178 CTGCATACCGCTACGCCAGCCAGCCCTACCCCTGCCACTGCCCTACAGTGACAGTTACG 237
QY 953 GACGAGTTTATGCTGCGGAGCCCTTACCACACACACTTGTCTCCAGCCCCACCTACGGCG 1012
Db 238 GACGAGTTTATGCTGCGGAGCCCTTACCACACGCACTTGTCTCCAGCCCCACCTACGGCG 297
QY 1013 TTGGTCCCATGAATGCTTTTGGCCCTTCCGAGCTTCCGAGTCCCAAGACTAGAGCCATGCTGATG 1072
Db 298 TTGGTCCCATGAATGCTTTTGGCACTTTGACCTTTGATGATGATGATGATGATGATGATG 357
QY 1073 ATGTGGGTCTCGTCTTTCTTCAATGCGAGCTAGTATATACCAAGGGGATACAAACCGTT 1132
Db 358 ATGTGGGTCTCGTCTTTCTTCAATGCGAGCTAGTATATACCGAGGGGATACAAACCGCT 417
QY 1133 TTGCTCCATATTAAATGATATAAAACCAATTAACAAACGCAAAACAAACAAACAA 1192
Db 418 TTGCTCCATATTAAATGATATAAAACCAATTAACAAACGCAAAACAAACAAACAA 447
QY 1193 AAAAACCAACCTTCCAAATGTTGGGAGAGAGAAAGCTTTCCGAGCCCGAGTGTTCGACA 1252
Db 448 -----AACCTTCCAATGTGGGAGAGAAAGCTNTCCGAGGCCCTGAGTATTGCAATA 500
QY 1253 CATGCGTAGGACATCATCTTTAGCAACTCAAGNACAAACGCAAAACAA 1299
Db 501 CATGCGAGGATGATCATCTTTTAGCAACTCTAAACAAACGCAAAACAAACAA 547

RESULT 13
AA859573 450 bp mRNA linear EST 03-JUL-1999
LOCUS UI-R-E0-bv-c-04-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
DEFINITION UI-R-E0-bv-c-04-0-UI 3' similar to emb|AL009266|HSP38A20 Homo
sapiens cDNA similar to RNA binding protein C. elegans, complete,
mRNA sequence.
ACCESSION AA859573
VERSION AA859573.1 GI:4230119
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
AUTHORS Fukuyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 460)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalizaition and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949093.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found. Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LILNL (info@image.lilnl.gov). IMAGE
ID=1777166
Seq primer: M13 Forward
POLYA=No.
Location/Qualifiers
1. .460
/organism="Rattus norvegicus"
```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 495)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 487.

FEATURES
source
1..495
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2244038"
/clone_lib="NCI CGAP GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
109 a 150 c 128 g 107 t 1 others

Query Match 29.7%; Score 398; DB 9; Length 495;
Best Local Similarity 87.2%; Pred. No. 1.3e-61;
Matches 463; Conservative 0; Mismatches 31; Indels 37; Gaps 1;

Qy 739 GCAGCTGATACCGAGGGGCTCAGCTTCGAGCGCGTGTGCGGAGTAGTATCAAGAGCCA 798
Db 1 GCGGCCGCTTACCGAGGGGCGACCTGCGAGCGCGCGGTGCGACCGTGTACACACCTTC 60

Qy 799 AGAGTGGCGGCGCCCAACCCCAATCCCGGCTATGCGGAGTAGTATCAAGAGCCA 858
Db 61 AGGGCGGGGCGCCCGCCCGATCCCGGCTACGCGGAGTAGTATCAAGAGCCT 120

Qy 859 GTGTATGCAATAAATTCCTACAGGGTGGTACGGTGTACCGTACCGTACGCCAGCCAC 918
Db 121 GTGTATGCAATAAATTCCTACAGGGTGGTATGCTGTGATACCGTACCGCAGCCTACC 180

Qy 919 CTGCGCTGTGCTGCTACAGTGACAGTACGACGAGTTATGCTGCGGACCCCTAC 978
Db 181 CTGCGCTGTGCTGCTACAGTGACAGTACGACGAGTTATGCTGCGGACCCCTAC 240

Qy 979 CACCACACACTTGTCTCCAGCCCCCACTACGCGGTGTGTCGATGAATGCTTTTGGCCCC 1038
Db 241 CACCACGCACTTGTCTCCAGCCCCCACTACGCGGTGTGTCGATGAATGCTTTTGCACCT 300

Qy 1039 TTGACCGATGCAAGATAGAGGCCATGCTGATGATGCGGTCTCGTCTTCTTCATTTG 1098
Db 301 TTGACTGATGCAAGATAGAGGCCATGCTGATGATGCGGTCTCGTCTTCTTCATTTG 360

Qy 1099 CAGGCTAGTATATACAGGGGGATACACCGTTTGTCTCCATATTAATGATTAACCA 1158
Db 361 CAGGCTAGTATATACAGGGGGATACACCGTTTGTCTCCATATTAATGATTAACCA 420

Qy 1159 TTAACAAACAGCAAAACAAACAAACAAACAAACAAACAAACCTTCCAAATGTGGGAG 1218
Db 421 TAAA-----AACCTTCCAAATGTGGGAG 443

Qy 1219 AGAGGAAGCTTTCCGAGGCCCGAGTGTTCGCACATGCAGTAGGACATCA 1269
Db 444 AAAGGAAGCTTTCGAGGGCTGAGTATTCGATACATGCAGTAGTACATCA 494

RESULT 11
AV729057 774 bp mRNA linear EST 17-OCT-2000
LOCUS AV729057 HTC Homo sapiens cDNA clone HTCBBG01 5', mRNA sequence.
DEFINITION AV729057
ACCESSION AV729057
VERSION AV729057.1 GI:10838478
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 774)

REFERENCE Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@hgsc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..774
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBBG01"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 186 a 217 c 180 g 191 t
ORIGIN
Query Match 29.2%; Score 391; DB 10; Length 774;
Best Local Similarity 83.0%; Pred. No. 2e-60;
Matches 487; Conservative 0; Mismatches 60; Indels 40; Gaps 2;

Qy 713 GCTTTCATATCCGGCCGACCTGCTCAGCTGCATACCGAGGGGCTCACCTTCGAGGCC 772
Db 1 GCTTTCGCTATCCAGCAGCACCTGCGGCCGCTACCGAGGGGCGCACCTTCGAGGCC 60

Qy 773 GTGGTCGACCCCTGTACAAACACCTTCAGAGCTGCGGCGCCCGCCCAATCCCGGCT 832
Db 61 GCGGTGCGACCTGTACAAACACCTTCAGGCGCGCGGCGCCCGCCCGATCCCGGCT 120

Qy 833 ATGGCGGAGTATGTATCAAGAGCCAGTGTATGGCAATAATTCGTACAGGGTGGTTACG 892
Db 121 ACGGCGGTGTGTATCCAGGATGGATTTATGG---TGCAGACATTTATGGTGGTTATG 177

Qy 893 CTGCATACCGCTAGCCCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 952
Db 178 CTGCATACCGCTAGCCCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 237

Qy 953 GACGAGTTATATCTGCGGACCCCTACCAACACACACTTGTCTCCAGCCCCCACTACGGG 1012
Db 238 GACGAGTTATATCTGCGGACCCCTACCAACACACACTTGTCTCCAGCCCCCACTACGGG 297

Matches 581; Conservative 0; Mismatches 76; Indels 15; Gaps 12;

QY 1 GCGCGCGCCCTGACACAAATGGCTCAG-CTTATGCTCTCAGCGAGTTGCTCCACCCCGAG 59
|||
Db 241 GCGGCTGCCCTGACACAAATGGCTCAGCTTACGCTTCGGCCCGAGTTGCTCCCGCGAG 300
|||

QY 60 AATGGCATCTT-CGAGAATACAGGCGCCCTCATCCCATCCCGGCGCAGAG-ACACCGGC 117
|||
Db 301 AACGGTATCCCGCGGAATACAGGCGCCCTCATCCCATCCCGGCGCAGAGTACACAGGC 360
|||

QY 118 CAGCCCACTGTCCTCCG-CACACATTAATCTTG-ATCTCTCTACAGAGCGACTCGGAG 175
|||
Db 361 CAGACACGGTTCCTCCGAGCACAATTAACTGTACCTCTCCCGCCGAGCGACTCCGAG 420
|||

QY 176 CAGAGC---GCTTATACAGCGCACAG-CGGTCTCCGGCACCCGCCACACAGACAGATGAT 231
|||
Db 421 CAGAGCCCGGCGACAGCGCTCAGACCGTCTCTGGCACCCGCCACACAGACAGATGAC 480
|||

QY 232 GCGGCCCGACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 291
|||
Db 481 CGAGCACCGGAGCGGATGGCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
|||

QY 292 CAGCCCAAGCGGCTGATGTCTCAACATCCCTTCCGGTTCGGGATCCGACCTCCGA 351
|||
Db 541 CAGCCCAAGCGGCTGATGTCTCAATATCCCTTCCGGTTCGGGATCCGAGCTCCGA 600
|||

QY 352 CAAATGTTGGCCAAATTTGGTAAATATTAGATGTTGAAATATTATTTTAATGAGCGGGC 411
|||
Db 601 CAAATGTTGGTCAATTTGGTAAATATTAGATGTTGAAATATTATTTTAATGAGCGAGC 659
|||

QY 412 TCGAAGGATTTGGTTTCTGTAATCTTCTGAAATAGTGGGATCGGACAGCGGCGAGGAG 471
|||
Db 660 TCAAGGATTTGGTTTCTGTAATCTTCTGAAATAGTGGGATCGGACAGCGGCGAGGAG 719
|||

QY 472 AATTCACGGTACCTGCTAGAGCGCGGTAAATTCGAGGTTAATATCGGACAGCACGC 531
|||
Db 720 AAA-TACACGGCACCTGCTAGAGCGCGGTAAATTCGAGGTTAATATCGGACAGCACGT 778
|||

QY 532 GTGATGACTAAATAAAGCGCGTGAACCCCTCACCAATGGCTGGAAATTAATATCA-GT 590
|||
Db 779 GTAATGACAAATAAAGACCGCTCAACCTTATACAAATGGCTGGAAATTAATTCAGT 838
|||

QY 591 TGTGGCGGGTCTACAGCCCG-ATTCTATGACGGCACGGTGC--TGTGTGCGAGGC 647
|||
Db 839 TGTGGTGCAGTCTACAGTCCCGAAATTTCTATGACGGCACGGTCTCTGTTGCGCAGCN 898
|||

QY 648 CACACAGGAGG 659
|||
Db 899 CCAACCGAGGAG 910
|||

RESULT 9
AL193117/c
LOCUS
DEFINITION
q95h03.x1 Soares fetal lung NBHL19W Homo sapiens cDNA clone
IMAGE:1744277 3' similar to TR:Q10572 Q10572 44.4 KD RNA-BINDING
PROTEIN IN FOX-1 LOCUS. ; mRNA sequence.

ACCESSION
AL193117
VERSION
AL193117.1 GI:3744326
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 457)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 895 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1..457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1744277"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGAGCGCGCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W." 132 c

BASE COUNT 110 a 140 c 105 g 132 t

Query Match 29.8%; Score 399.4; DB 9; Length 457;
Best Local Similarity 92.1%; Pred. No. 7.6e-62;
Matches 421; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 272 AAAACACAGAAACAAAGTCCAGCCCAAGCGCTGCATGTGTCCAAACATCCCTCCCGT 331
|||
Db 457 AAAACACAGAAACAAAGTCTCAGCCCAAGCGCTGCATGTCTCCAAATATCCCTCCAGT 398
|||

QY 332 TCCGGATCCAGACCTTCGACAAATGTTTGGCCAAATTTGGTAAATATTAGATGTTGAAA 391
|||
Db 397 TCCGGATCCGGACCTCAGACAAATGTTTGGTCAATTTGGTAAATCTTTAGATGTTGAAA 338
|||

QY 392 TTATTTTAAATCAGCGGGCTCGAAGGATTTGGTTGTAACCTTCGAAATAGTGGCG 451
|||
Db 337 TTATTTTAAATCAGCGAGGCTCAAAGGATTTGGTTGTAACCTTCGAAATAGTGGCG 278
|||

QY 452 ATCGGACAGCGCGAGGAGAAATTCACGTCACGTCGTGTAGAGCGCGTAAATTCGAG 511
|||
Db 277 ATCGGACAGCGCGAGGAGAAATTCACGTCACGTCGTGTAGAGCGCGTAAATTCGAG 218
|||

QY 512 TTAATAATGCGACAGCAGCGGTGATGACTAATAAAAGCGCGTGAACCCCTACACCAATG 571
|||
Db 217 TAAATAATGCGACAGCAGCGGTGATGACTAATAAAAGCGCGTGAACCCCTTATACAAATG 158
|||

QY 572 GCTGAAATTAATCCAGTTGTGGCGGCTCTACAGCCCGGACTTCTATGAGGCGCGG 631
|||
Db 157 GCTGAAATTAATCCAGTTGTGGCGGCTCTACAGTTCCAGTCCCGAATTTCTATGAGGCGCGG 98
|||

QY 632 TGCTGTTGTGCCAGCGCAACAGAGGAGGATCTTCCATGTACAGTGGCCCCCAGTTTCACTTG 691
|||
Db 97 TCCTGTTGTGCCAGCGCAACAGAGGAGGATCTTCCATGTACAGTGGCCCCCAGTTTCACTTG 38
|||

QY 692 TATATACTTCTGCAATGCTGCTTCCATATTCGCGC 728
|||
Db 37 TATATACTTCTGCAATGCGAGGCTTCCCGGTATCCAGC 1
|||

RESULT 10
AL65926
LOCUS
DEFINITION
A165926
495 bp mRNA linear EST 04-MAY-1999
tt48e08.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2244038 3'
similar to TR:O43251 O43251 HYPOTHETICAL 39.5 KD PROTEIN. ; mRNA
sequence.
ACCESSION
A165926
VERSION
A165926.1 GI:4740905
KEYWORDS
EST.
SOURCE
human.

```

Db 301 CCCAAGCGGCTGCATGTGTCAACATCCCCCTTCGCGTCCGGGATCCAGACCTCCGACAA 360
Qy 355 ATGTTTGGCCAAATTTGTAATAATTAGATGTTGAAATTAATTTTAATAGCGGGGCTCG 414
Db 361 ATGTTTGGCCAAATTTGTAATAATTAGATGTTGAAATTAATTTTAATAGCGGGGCTCG 420
Qy 415 AAGGATTTGGTTTCGTAACCTTTCGAAAAATAGTCGGATCGCGACAGGGGAGGAGAAA 474
Db 421 AAGGATTTGGTTTCGTAACCTTTCGAAAAATAGTCGGATCGCGACAGGGGAGGAGAAA 480
Qy 475 TTGCACGGTACCGT 488
Db 481 TTGCACGGTACCGT 494

RESULT 7
AI291784/c
LOCUS      AI291784               479 bp      mRNA      linear      EST 29-JAN-1999
DEFINITION qm7508.x1 Soares_placenta 8to9weeks 2NbHP8to9W Homo sapiens cDNA
            clone IMAGE:1894551 3' similar to TR:Q10572 Q10572 44.4 KD
            RNA-BINDING PROTEIN IN FOX-1 LOCUS. ;, mRNA sequence.
ACCESSION  AI291784
VERSION     AI291784.1
KEYWORDS   AI291784.1 GI:3934558
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 479)
            NCI-CoAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 863 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 452.
FEATURES   source
            1..479
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1894551"
            /clone_lib="Soares_placenta 8to9weeks 2NbHP8to9W"
            /dev_stage="two placentae: One from 8 weeks and another
            from 9 weeks post conception"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Organ: placenta; Vector: p7T3D (Pharmacia) with a
            modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGCGCGATTTTCTTTTCTTTT 3'],
            double-stranded cDNA was size selected, ligated to Eco RI
            adapters (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of a modified p7T3 vector
            (Pharmacia). Library constructed by Bento Soares and
            M. Fatima Bonaldo."
BASE COUNT 112 a 118 c 115 g 134 t
ORIGIN
Query Match 30.9%; Score 413.4; DB 9; Length 479;
Best Local Similarity 91.4%; Pred. No. 2.3e-64;
Matches 438; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 268 TCTGAAAAACACAGAAAAAAGTCCAGCCCGGCTGCATGTGTCCAAATCCCTTC 327
Db 479 TCTGAAAAACACAGAAAAAAGTCTCAGCCAGCGCTGCATGTCTCCAAATATCCCTTC 420
Qy 328 CGGTTCCGGATCCAGACTCCGCAAAATGTTGGCCAAATTTGGTAAATATAGATGTT 387
Db 419 AGGTTCCGGATCCCGGACCTCAGACAAATGTTTGGTCAATTTGGTAAATATCTAGATGTT 360

```

```

Qy 388 GAAATTATTTTAAATGACGGGGCTCGAAGGGATTTGGTTTCGTAACCTTTCGAAAAATAGT 447
Db 359 GAAATTATTTTAAATGACGGAGGCTCAAGGGGATTTGGTTTCGTAACCTTTCGAAAAATAGT 300
Qy 448 GCCGATTCGCGNACAGGCGAGGGAGAAATTCACCGGTACCGTGTAGAGGGCCGCTAAATC 507
Db 299 GCCGATTCGCGNACAGGCGAGGGAGAAATTCACCGCACCGTGTAGAGGGCCGCTAAATC 240
Qy 508 GAGGTTAATTAATCGCAGACGACCGGTGATGACTAATAAAAAAGCCGCTGAACCCCTACACC 567
Db 239 GAGGTTAATTAATGCCACAGCACCGTGTAAATGACAAATAAAAAAGACCCGCTCAACCCCTTATACA 180
Qy 568 AATGGCTGGAAATTAATCCAGTTGTGGCGCGTCTACAGCCCGCACTTCTATGACAGGC 627
Db 179 AATGGCTGGAAATTTGAATCCAGTTGTGGGTGCAGTCTACAGTCCCGAATTTCTATGACAGGC 120
Qy 628 ACGGTGCTGTGTGGCAGGCGCAACACGAGGGGATCTTCCATGTACAGTGCCCGCAGTTCA 687
Db 119 ACGGTCTGTGTGGCAGGCGCAACACGAGGGGATCTTCCATGTACAGTGCCCGCAGTTCA 60
Qy 688 CTTGTATATATCTTCTGCAATGCGTGTTCCTTCCATATCCGGCGCCCATCTGTCAGCTGC 746
Db 59 CTTGTATATATCTTCTGCAATGCCAGGCTTCACGTATCCAGCAGCCACCGCGCGCGCGC 1

RESULT 8
BM808707
LOCUS      BM808707               1020 bp      mRNA      linear      EST 05-MAR-2002
DEFINITION AGENCOURT 6618011 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5734270
            5', mRNA sequence.
ACCESSION  BM808707
VERSION     BM808707.1
KEYWORDS   BM808707.1 GI:19125530
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1020)
            NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L2AM12739 row: e column: 23
            High quality sequence stop: 723.
FEATURES   source
            1..1020
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5734270"
            /clone_lib="NIH MGC 124"
            /tissue_type="hippocampus"
            /lab_host="DH10B"
            /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
            (destroyed); Site 2: NotI; RNA source male hippocampus,
            age 27. Library is oligo-dT primed and directionally
            cloned (EcoRV site is destroyed upon cloning). Average
            insert size 1.4 kb, insert size range 0.9-4 kb. Library is
            normalized and enriched for full-length clones and was
            constructed by C. Gruber (Invitrogen). Research Genetics
            tracking code 012."
BASE COUNT 274 a 279 c 275 g 189 t
ORIGIN
Query Match 30.0%; Score 401.6; DB 14; Length 1020;
Best Local Similarity 86.5%; Pred. No. 2.3e-62;

```

cdna Library Preparation: M.B. Soares Lab Clone Distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (infoimage.llnl.gov). IMAGE ID= 1777166
Seq primer: M13 Forward.

FEATURES

Location/Qualifiers
1..533
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

BASE COUNT 134 a 125 c 164 g 109 t 1 others
ORIGIN

Query Match 33.2%; Score 444.4; DB 12; Length 533;
Best Local Similarity 93.0%; Pred. No. 6.5e-70;
Matches 492; Conservative 0; Mismatches 22; Indels 15; Gaps 2;
QY 648 CAACAGGAGGATCTCCATGTACAGTGGCCCGAGTTCACCTGTATATACCTTCTGCAA- 706
Db 533 CAACAGGAGGATCTCCATGTACAGTGGCCCGAGTTCACCTGTATATACCTTCTGCAA 474
QY 707 -----TCCCTGGCTTCCATATCCGCGCGCCACCTGTCAGCTGCATACCGAGG 755
Db 473 TTCTGTTCCAGTCCCTGGCTTCCATATCCGCGCGCCACCTGTCAGCTGCATACCGAGG 414
QY 756 GGCTACACCTTCGAGGCGGTGGTGCACCGTGTACAAACCTTCAGAGCTGCGCGCGCCCC 815
Db 413 GGCTACACCTTCGAGGCGGTGGTGCACCGTGTACAAACCTTCAGAGCTGCGCGCGCCCC 354
QY 816 ACCCCCAATCCCGCTATGCGGAGTAGTGTATCAAGAGCAGTGTATGGCAATAAAT 875
Db 353 ACCCCCAATCCCGCTATGCGGAGTGTGTTTACCAGGATGGATTTTATGG---TGCAGA 297
QY 876 GCTACAGGTGTTACGCTGCATACCGCTACGCCAGCCAGCCACCCCTGCCACTGCTGCTC 925
Db 296 CATTTATGGTGTTCGCTGCATACCGTACGCCAGCCAGCCACCCCTGCCACTGCTGCTC 237
QY 936 CTACAGTGACAGTTACGAGCAGTATTATGCTGCGACCCCTACCCACACACACTTGTCTCC 995
Db 236 CTACAGTGACAGTTACGAGCAGTATTATGCTGCGACCCCTACCCACACACACTTGTCTC 177
QY 996 AGCCCCCACTACGCGGTGGTGCATGAATGCTTTTGGCCCTTACCGCATGCCAAGAC 1055
Db 176 AGCCCCCACTACGCGGTGGTGCATGAATGCTTTTGGCCCTTACCGCATGCCAAGAC 117
QY 1056 TAGGAGCCATGCTGATGATGCTGGTTCGTTCTTCTTCAATTCAGGCTAGTATATACCA 1115
Db 116 TAGGAGCCATGCTGATGATGCTGGTTCGTTCTTCTTCAATTCAGGCTAGTATATACCG 57
QY 1116 AGGGGATACCAACCGTTTGTCTCCATATTAAATGATAAAACCAATTAAC 1164
Db 56 AGGGGATACCAACCGTTTGTCTCCATATTAAATGATAAAACCAATTAAC 8

RESULT 6
BF420112 494 bp mRNA linear EST 28-NOV-2000
LOCUS UI-R-BJ2-bpj-a-08-0-UI.s1 UI-R-BJ2 Rattus norvegicus cdna clone
DEFINITION UI-R-BJ2-bpj-a-08-0-UI 3', mRNA sequence.
ACCESSION BF420112
VERSION BF420112.1 GI:11408101

KEYWORDS
SOURCE
ORGANISM

EST.
Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus

1 (bases 1 to 494)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Oligo-dT track not found, Not I site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab Clone Distribution: clones will be available through

Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=No.

FEATURES
source

Location/Qualifiers

1..494

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-BJ2-bpj-a-08-0-UI"

/clone_lib="UI-R-BJ2"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ2

library is a subtracted library derived from the following

tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,

atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15

dpc, AV canal at 15 dpc. For a detailed description of

the library from which this clone was derived, please

visit our web site at rategen.uiowa.edu. The

subtraction has been previously described in (Bonaldo,

Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_SEQ=None found"

BASE COUNT 128 a 162 c 115 g 89 t

ORIGIN

Query Match 31.1%; Score 417.2; DB 12; Length 494;
Best Local Similarity 98.2%; Pred. No. 4.9e-65;
Matches 485; Conservative 0; Mismatches 3; Indels 6; Gaps 6;

QY 1 GCGGCGCCCTGACACAATGGCTCAG-CTTATGCTCAGCGCAGTTGCTCCACCCCGAG 59

Db 1 GCGGCGCCCTGACACAATGGCTCAGCCTTATGCTCAGCGCAGTTGCTCCACCCCGAG 60

QY 60 AATGGCAT-CCTGCGAGATACACGGCCCTCATCCCATCCCGCGCCAGAG-ACACCGGC 117

Db 61 AATGGCATCCTGCAGATACACGGCCCTCATCCCATCCCGCGCCAGAGTACACCGGC 120

QY 118 CAGCCCACTGTCCCG-CCACACATTAAACTTGT-ATCTCTTACACAGACGCACTCGGAG 175

Db 121 CAGACCACTGTCCCGACCACTTAAACCTGTATCTCTCTACACAGACGCACTCGGAG 180

QY 176 CAGAGGCTTATACAGCGCAGAG-CCGTCTCCGCGCCCGCCACACAGACAGATGATGCC 234

Db 181 CAGAGGCTGATACAGCGCAGACCGTCTCCGCGCCCGCCACACAGACAGATGATGCC 240

QY 235 GCGCCGACCGCGCGCCAGCCCGCCAGACAAACCTTCTGAAAAACACAGAAACAGTCCCGAG 294

Db 241 GCGCCGACCGCGCGCCAGCCCGCCAGACAAACCTTCTGAAAAACACAGAAACAGTCCCGAG 300

QY 295 CCAAGCGGCTGATGTGTCCAAACATCCCTTCCGTTCCGGATCCAGACCTCCGACAA 354

ORIGIN

QY 302 GGCTGATGTGTCACATCCCTTCCTCGGTTCGGGATCCAGACCTCCGACAAATGTTTG 361
 Db 301 GGCTGATGTGTCACATCCCTTCCTCGGTTCGGGATCCAGACCTCCGACAAATGTTTG 360
 QY 362 GCCAATTTGGTAAATATAGATGTTGAAATATATTTTAAATGAGCGGGCTCGAAGGAT 421
 Db 361 GCCAATTTGGTAAATATAGATGTTGAAATATATTTTAAATGAGCGGGCTCGAAGGAT 420
 QY 422 TTGGTTTCGTAACTTTTCGAAATAGTTCGGGATCGGACAGAGCGGAGGAGAAATTTGCACG 481
 Db 421 TTGGTTTCGTAACTTTTCGAAATAGTTCGGGATCGGACAGAGCGGAGGAGAAATTTGCACG 480
 QY 482 GTACCGTGTAGAGCGGCGTAAATTCGAGGTTTAATTAATCGACAGCACGCGTGATGACTA 541
 Db 481 GTACCGTGTAGAGCGGCGTAAATTCGAGGTTTAATTAATCGACAGCACGCGTGATGACTA 540
 QY 542 AT 543
 Db 541 AT 542
 RESULT 4
 A1885361/c
 LOCUS
 DEFINITION w192h03.x1 NCI CGAP Brn25 Homo sapiens cDNA clone IMAGE:2432405 3' similar to TR:O43251 O43251 HYPOTHETICAL 39.5 KD PROTEIN. ;, mRNA sequence.
 A1885361
 A1885361.1 GI:5590525
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 587)
 TITLE NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CCAP/BTGP), Tumor Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 728 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 459.
 Location/Qualifiers
 1..587
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI CGAP Brn25"
 /tissue_type="anaplastic oligodendroglioma"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGATAGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
 121 a 143 c 156 g 167 t

BASE COUNT

Query Match 34.6%; Score 464.2; DB 9; Length 587;
 Best Local Similarity 90.1%; Pred. No. 1.8e-73;
 Matches 520; Conservative 0; Mismatches 53; Indels 4; Gaps 2;

QY 141 TTAACCTTGATCTCTCTACACAGACGCACTCGGAGCAGAGC---GCTTATACAGCCGAC 197
 Db 577 TAAACCTTGATCTCTCTACACAGACGCACTCGGAGCAGAGC---GCTTATACAGCCGAC 518
 QY 198 AG-CCGCTCTCGGACCGCCACACAGACATGATGCGCCCGCCAGCCAGCCAGCCGCCCC 256
 Db 517 AGACCGTCTCTGGACCGCCACACAGACATGATGCGCCCGCCAGCCAGCCAGCCGCCCC 458
 QY 257 AGACACAACCTTCTGAAAAACACAGAAACAAAGTCCAGAGCCCAAGCGCTGATGTGCTCA 316
 Db 457 AGACACAACCTTCTGAAAAACACAGGAAACAAAGTCTCAGCCCAAGCGCTGATGTCTCCA 398
 QY 317 ACATCCCTTTCGGTTCGGGATCCAGACCTCCGACAAATGTTTGGCCCAATTTGGTAAAA 376
 Db 397 ATATCCCTTTCAGGTTCGGGATCCGACCTCCAGACAAATGTTTGGTCAATTTGGTAAAA 338
 QY 377 TATTAGATGTTGAAATTTATTTTAAATGAGCGGGCTCGAAGGGATTTGGTTTCGTAACCTT 436
 Db 337 TCTTAGATGTTGAAATTTATTTTAAATGAGCGAGCTCAAGGGATTTGGTTTCGTAACCTT 278
 QY 437 TCGAAAAATAGTTCGGATCGGACAGGCGAGGGAGAAATTCACCGGTACCGTGTGAGG 496
 Db 277 TCGAAAAATAGTTCGGATCGGACAGGCGAGGGAGAAATTCACCGGTACCGTGTGAGG 218
 QY 497 GCCGTAATAATCGAGTTAATAATCGGACAGCAGCGGTGATGACTAATAAAGAGCCGTGA 556
 Db 217 GCCGTAATAATCGAGTTAATAATCGGACAGCAGCGGTGATGACTAATAAAGAGCCGTGA 158
 QY 557 ACCCTTACACCAATGCGTGGAAATTAATCCAGTTGTTGGCGCGGTCTACAGCCCGGACT 616
 Db 157 ACCCTTATACAAATGCGTGGAAATTTGAATCCAGTTGTTGGTGCAGTCTACAGTCCCGAAT 98
 QY 617 TCTATGAGGACACGGTGTGTTGTCAGGCGCAACACGAGGGGATCTTCCATGTACAGTG 676
 Db 97 TCTATGAGGACACGGTGTGTTGTCAGGCGCAACACGAGGGGATCTTCCATGTACAGTG 38
 QY 677 GCCCAGTTTCACTTGTATATACCTTCTGCAATGCCTGG 713
 Db 37 CCCCCAGTTTCACTTGTATATACCTTCTGCAATGCCTGG 1

RESULT 5

BF549922/c

LOCUS

DEFINITION

BF549922

UI-R-E0-bv-C-04-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone

BF549922

BF549922.1 GI:11659652

EST.

KEYWORDS

SOURCE

ORGANISM

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 533)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

MEDLINE

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Contact: nisc_mgc@nhgri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantrilpop, S., Thomas, P.J., Touchman, J.W., Tsurgelon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 79 Row: j Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922072
This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers
1. 973
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5734270"
/tissue type="Brain, hippocampus"
/clone_lib="NIH MGC_124"
/lab_hosts="DHIOB"
/notes="Vector: pCMV-SPORT6"

BASE COUNT 265 a 270 c 256 g 182 t

ORIGIN

Query Match 39.5%; Score 529.2; DB 11; Length 973;
Best Local Similarity 88.9%; Pred. No. 3.7e-85;
Matches 663; Conservative 0; Mismatches 73; Indels 10; Gaps 8;

QY 1 GCGGCCGCTGTACACAAATGGCTCAG-CTTATGCTCAGCGCAGTTTCGTCTCCACCCCG 59
DB 228 GCCGCTGCCCTGTACACAAATGGCTCAGCCTTACGCTTCGGCCAGTTTGTCTCCCGCAG 287

QY 60 AATGGCATCTT-GCAGAAATACACGGCCCTCATCCCCATCCCGCCGCGAGAG-ACACCGGC 117
DB 288 AACGGTATCCCGGGGAATACACGGCCCTCATCCCCACCGCCGCGAGATACACAGGC 347

QY 118 CAGCCCACTGTCCCG-CCACACATTAACCTTG-ATCCTCTACACAGCGCACTCCGAG 175
DB 348 CAGACCAGGTTCCCGAGCACACATTAACCTGTACCTCTCCGCCACAGCGCACTCCGAG 407

QY 176 CAGAGC--GCTTATACAAGGGCAGAC-CCGTCTCCGCGACCGCCACACAGACAGATGAT 231
DB 408 CAGAGCCGGGGACAGAGCGCTCAGACCGCTCTCTGGCCCGCCACACAGACAGATGAC 467

QY 232 GCCGCCCGGACCGGCGCCGCGCCAGACACACCTTTCTGAAAAACACAGAAAAACAAGTCC 291
DB 468 GCAGCAGCGGATGGCCAGCCCGCAGACACACCTTCTGAAAAACACGGAACAAGTCT 527

QY 292 CAGCCCAAGCGGCTGCATGTGTCAACATCCCTTCGGTTCGGGATCCAGACTCCGA 351
DB 528 CAGCCCAAGCGGCTGCATGTGTCTCAATATCCCTTCAGGTTCCGGGATCCGGACCTCAGA 587

QY 352 CAATGTTTGGCCAAATTTGGTAAATATATAGATGTTGAAATATTTTAAATGAGCGGGC 411
DB 588 CAATGTTTGGTCAATTTGGTAAATATTTAGATGTTGAAATATTTT-AATGAGGAGGC 646

QY 412 TCGAAGGATTTGGTTTCGTAACCTTTTCGAAAAATAGTCGGATCGGACAGGCGGAGGAG 471
DB 647 TCAAGGATTTGGTTTCGTAACCTTTTCGAAAAATAGTCGGATCGGACAGGCGGAGGAG 706

QY 472 AAATTGACCGGTACCGGTAGAGGGCGGTAAATATCGAGGTTAAATATGCGACAGACGCC 531
DB 707 AAATTACAGCGCACCGGTAGAGGGCGGTAAATATCGAGGTTAAATATGCGACAGACGCT 766

QY 532 GTGATGACTAATAAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATTAATCCAGTT 591
DB 767 GTAATGACAAATAAAGACCGGTCAACCCCTTATCAATATGCTGGAATTAATTAATCCAGTT 826

QY 592 GTGGCGCGGTCTACAGCCCGCACTTCTATGACGCGCGGTGCTGTGTGTGCGAGGCCAAC 651

Db 827 GTGGGTGCACTTACAGTCCCGAATTTCTATGAGGACGGTCTGTGTGCCAGGCCAAC 886
QY 652 CAGAGGGATTTCCCATGTACAGTGGCCCCAGTTCCTTGTATATATCTTCTGCAATGCCCT 711
DB 887 CAGAGGGATTTCCCATGTACAGTGGCCCCAGTTCCTTGTATATATCTTCTGCAATGCCA 946

QY 712 GGTTTCCATATCCGGCCGCACTGC 737
DB 947 GGTTCCCGTATCCAGCAGCCACCGC 972

RESULT 3
LOCUS AI101278
DEFINITION EST210567 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBRBK70 3' end, mRNA sequence.
ACCESSION AI101278
VERSION AI101278.1 GI:3704621
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 542)
AUTHORS Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
Location/Qualifiers
1. 542
/organism="Rattus sp."
/db_xref="ATCC (inhost):2022664"
/db_xref="taxon:10118"
/clone="RBRBK70"
/clone_lib="Normalized rat brain, Bento Soares"
/note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 146 a 168 c 128 g 100 t
ORIGIN

Query Match 34.7%; Score 465.2; DB 9; Length 542;
Best Local Similarity 98.3%; Pred. No. 1.2e-73;
Matches 533; Conservative 0; Mismatches 3; Indels 6; Gaps 6;

QY 8 CCCTTGACCAATGGTTCAG-CTTATGCTCAGCGAGTTTCGTCCACCCAGAAATGGCA 66
DB 1 CCCTTGACCAATGGTTCAGCCTTATGCTTCAGCGAGTTTCGTCCACCCAGAAATGGCA 60

QY 67 T-CCTGAGAAATACAGGCGCCCTCATCCCATCCCGCCGAGAG-ACACCGGCGAGCCCA 124
DB 61 TCCTGAGAAATACAGGCGCCCTCATCCCATCCCGCCGAGAGTACACCGGCGAGACCA 120

QY 125 CTGTCCCG-CCACACATTAACCTTG-ATCCTCTACACAGCGCACTCGGAGCAGAGCG 182
DB 121 CTGTCCCGACCAACATTAACCTGTATCTCTACAGAGCGCACTCGGAGCAGAGCG 180

QY 183 CTTATACAAGCGCACAG-CCGTCTCCGGCACCCGCCACACAGACAGATGATGCCGCCCGCA 241
DB 181 CTGATACAAGCGCACAGCGCTCTCCGGCACCCGCCACACAGACAGATGATGCCGCCCGCA 240

QY 242 CCGACGGCGAGCCCCAGACACAACTTCTGAAAAACACAGAAAAACAAGTCCCGCCCGAGC 301
DB 241 CCGACGGCGAGCCCCAGACACAACTTCTGAAAAACACAGAAAAACAAGTCCCGCCCGAGC 300

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, J., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 32 Row: 1 Column: 9
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922072
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers

1..4039

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4815500"

/tissue type="Brain, hippocampus"

/clone lib="NIH MGC_95"

/lab host="DH10B"

/note="vector: pBluescript"

BASE COUNT 1201 a 858 c 876 g 1104 t

ORIGIN

Query Match 68.0%; Score 911.2; DB 11; Length 4039;

Best Local Similarity 85.9%; Pred. No. 1.2e-153;

Matches 1144; Conservative 0; Mismatches 138; Indels 50; Gaps 10;

```

Qy 1 GCGGCCGCCCTGACCAATGGCTCAG-CTTATGCTCAGCGCAGTTTCGCTCCACCCGAG 59
Db 280 GCCATGCCCCCTGACCAATGGCTCAGCCTTACGCTTCGGCCAGTTTGTCTCCCGCGAG 339

Qy 60 AATGCGATCTT-GCAGAATACAGCGCCCTCATCCCATCCGCGCCAGAG-ACACCGG 117
Db 340 AACGGTATCCCGCGGAATACACGCGCCCTCATCCCATCCCGCGCCAGAGTACACAGCG 399

Qy 118 CAGCCCACTGTCCCG-CCACACATTAACTTG-ATCTCTCTACACAGCAGCAGCTCGAG 175
Db 400 CAGACCAAGTTCGCGAGCACATTAACCTGTACCTTCGCGCCAGCAGCAGCTCCGAG 459

Qy 176 CAGAGC--GCTTATACAGCGCACAG-CCGTCTCCGCGCACCGCCACACAGACATGAT 231
Db 460 CAGAGCCCGCGGACACAGCGCTCAGACCGGTATCTGGCAGCCGACACACAGACATGAC 519

Qy 232 GCCGCCGACCGAGCGCCAGCCAGACACAACTTCTGTAACACACAGAAACAGTCC 291
Db 520 CAGACCGCGAGGATGGCCAGCGCCAGACACAACTTCTGTAACACACAGGAAACAGTCT 579

Qy 292 CAGCCCAAGCGCTGATGTGTCACATCTCCCTTCGCGTTCGCGATCCAGACCTCCGA 351
Db 580 CAGCCCAAGCGCTGATGTGTCACATCTCCCTTCGAGTTCGCGATCCGAGCTCCAGA 639

Qy 352 CAAATGTTTGGCCAAATTTGGTAAATATTAGATGTTGAAATATTATTTTAAAGCGGGC 411
Db 640 CAAATGTTTGGTCAATTTGGTAAATCTTAGATGTTGAAATATTATTTTAAAGCGAGGC 699

Qy 412 TCGAAGGATTTGGTTTGGTAAATATTAGATGTTGAAATATTAGATGTTGAAATATTAG 471
Db 700 TCGAAGGATTTGGTTTGGTAAATATTAGATGTTGAAATATTAGATGTTGAAATATTAG 759

Qy 472 AAATTCACCGGTACCGTGTAGAGGCGCGTAAATTCGAGGTTTAAATTCGACACAGCAGC 531
Db 760 AAATTCACCGGTACCGTGTAGAGGCGCGTAAATTCGAGGTTTAAATTCGACACAGCAGT 819

Qy 532 GTGATGACTTAATAAAGGCGGTGAACCCCTACACCAATGGCTGGAAATTAATCCAGTT 591
Db 820 GTAATGACAAATAAAGAGCCGTCAACCTTATACAAATGGCTGGAAATTAATCCAGTT 879

Qy 592 GTGGCGCGGTCTACAGCCCGACTCTATGAGGACCGGTCTGTTGTCAGGCGCAAC 651
Db 880 GTGGGTGACAGTCTACAGTCCCGAATTTCTATGAGGACCGGTCTGTTGTCAGGCGCAAC 939

Qy 652 CAGGAGGATCTTCAGATGACAGTGGCCCCAGTTCATCTGTATATATCTCTCAATGCT 711
Db 940 CAGGAGGATCTTCAGATGACAGTGGCCCCAGTTCATCTGTATATATCTCTCAATGCTCA 999

```

```

Qy 712 GCGTTTCCATATCCGGCGCGGCGGCTGCTGAGCTGCTACGAGGGGCTCACCTTCGAGGC 771
Db 1000 GCGTTTCCGATATCCAGCAGGCGCGCGCGCGCGCTTACCGAGGGGCGCACCTTCGAGGC 1059

Qy 772 CGTGGTGCACCGGTGTAACACACCTTCAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 831
Db 1060 CGCGGTGCACCGGTGTAACACACCTTCAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1119

Qy 832 TATGCGGAGTGTATCAAGAGCGCAGTGTATGGCAATAAATTTGCTACAGGGTGGTTAC 891
Db 1120 TACGGCGGTGTTGTTTACCAGGATGGATTTATGG---TGCAGACATTTATGTTGGTTAT 1176

Qy 892 GCTGCATACCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
Db 1177 GCTGCATACCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1236

Qy 952 GGAGGAGTTTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011
Db 1237 GGAGGAGTTTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295

Qy 1012 GTTGGTGCCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGGAGCCATCTGAT 1071
Db 1296 GTTGGTGCCATGAATGCTTTTGGCACCTTTGACTGATGCCAAGACTAGGAGCCATCTGAT 1355

Qy 1072 GATGTGGTCTCGTCTCTTCTTCTTCTGAGCGCTAGTATATACCAAGGGGATACACCGT 1131
Db 1356 GATGTGGTCTCGTCTCTTCTTCTTCTGAGCGCTAGTATATACCAAGGGGATACACCGT 1415

Qy 1132 TTTGCTCCATATTAATGATAAAACCATTAACAAACAAAGCAAAACAAACAAACAAACA 1191
Db 1416 TTTGCTCCATATCTAATGACAAACCAATAAA----- 1446

Qy 1192 AAAAAACCAACCTTCCAATGTGGGAGAGAGAGTTCCTCGAGCGCGGAGTGTGGCGAC 1251
Db 1447 -----AACCTTCCAATGTGGGAGAGAGAGTTCCTCGAGCGCGTGTGCAAT 1498

Qy 1252 ACATGCTAGTAGACATCATCTTTAGCACTCAAGAAACAAAGCAAAACAAACAAACAA 1311
Db 1499 ACATGCTAGTAGACATCATCTTTAGCACTCTAAAAACAAACAAACAAACAAACAAAG 1558

Qy 1312 AAAATAAGCGCG 1323
Db 1559 AAAAAAATGAC 1570

```

RESULT 2

BC033885

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC033885 Homo sapiens, clone IMAGE:5734270, mRNA.
 973 bp mRNA linear HTC 08-JUL-2002
 BC033885
 BC033885.1 GI:21707343
 HTC.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 973)
 Strausberg, R.
 Direct Submission
 Submitted (02-JUL-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: Invitrogen
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:40:33 ; Search time 1884.53 Seconds
(without alignments)

11515.871 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340

Sequence: 1 gggccgcgcctgacacaat.....ggccgaagggttcgctaga 1340

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hct:*

9: gb_est1:*

10: gb_est2:*

11: gb_hct:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estcom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	911.2	68.0	4039	11 BC026312	BC026312 Homo sapi
2	529.2	39.5	973	11 BC033885	BC033885 Homo sapi
3	465.2	34.7	542	9 AI101278	AI101278 EST210567
4	464.2	34.6	587	9 AI885361	AI885361 w192h03.x
5	444.4	33.2	533	12 BF549922	BF549922 UI-R-E0-b
6	417.2	31.1	494	12 BF420112	BF420112 UI-R-BJ2-

C	7	413.4	30.9	479	9	AI291784	AI291784 qm75b08.x
C	8	401.6	30.0	1020	14	BM808707	BM808707 AGENCOURT
C	9	399.4	29.8	457	9	AI193117	AI193117 q6e9h03.x
C	10	398	29.7	495	9	AI656926	AI656926 tt48e08.x
C	11	391	29.2	774	10	AV729057	AV729057 AV729057
C	12	384.2	28.7	773	10	AV729198	AV729198 AV729198
C	13	383.2	28.6	460	9	AA859573	AA859573 UI-R-E0-b
C	14	379.8	28.3	549	10	AV655025	AV655025 AV655025
C	15	378.2	28.2	451	9	AI193843	AI193843 q67h11.x
C	16	376.8	28.1	428	9	AA460940	AA460940 z261d04.s
C	17	372.8	27.8	449	9	AI133596	AI133596 qt42e01.x
C	18	372.6	27.8	428	10	AV665024	AV665024 AV665024
C	19	368	27.5	464	9	AI332614	AI332614 qp94h10.x
C	20	357.4	26.7	391	9	AI186273	AI186273 q20b08.x
C	21	345.4	25.8	480	9	AI401040	AI401040 th27a12.x
C	22	343.4	25.6	477	10	AW197589	AW197589 xm44g02.x
C	23	338.8	25.3	607	9	AU145000	AU145000 AU145000
C	24	337	25.1	379	9	AI843387	AI843387 UI-M-AQ1-
C	25	334	24.9	893	9	AU118435	AU118435 AU118435
C	26	332	24.8	565	14	W68285	W68285 z835e10.81
C	27	330.2	24.6	449	9	AI374891	AI374891 ts60g12.x
C	28	330.2	24.6	456	9	AI917298	AI917298 ts96g09.x
C	29	328.2	24.5	459	9	AI244212	AI244212 q186b11.x
C	30	328	24.5	430	9	AI850237	AI850237 UI-M-BG1-
C	31	316.6	23.6	395	9	AI717457	AI717457 UI-R-Y0-a
C	32	314	23.4	383	9	AI189407	AI189407 qd01h05.x
C	33	313.2	23.4	361	9	AI835011	AI835011 UI-M-AM1-
C	34	311.6	23.3	425	10	BB750925	BB750925 BB750925
C	35	311.2	23.2	723	12	BG306387	BG306387 fm58d11.x
C	36	301.4	22.5	443	9	AA773715	AA773715 af81c05.x
C	37	295.8	22.1	347	9	AI041587	AI041587 ox56g05.x
C	38	295.8	22.1	439	9	AI968176	AI968176 wul4a06.x
C	39	295.8	22.1	439	10	AW589795	AW589795 hg22d04.x
C	40	295.6	22.1	801	9	AU051767	AU051767 AU051767
C	41	295.2	22.0	1589	11	BC002124	BC002124 MUB MUSCU
C	42	294.8	22.0	439	10	BE501380	BE501380 fv41h09.x
C	43	288.2	21.5	521	13	BM183349	BM183349 fv10e03.x
C	44	280	20.9	420	9	AI655094	AI655094 wb57a10.x
C	45	279.2	20.8	538	9	AA818804	AA818804 UI-R-A0-a

ALIGNMENTS

RESULT 1	BC026312	4039 bp	mRNA	linear	HTC 08-APR-2002
LOCUS	BC026312	Homo sapiens, clone IMAGE:4815500, mRNA.			
DEFINITION	BC026312				
ACCESSION	BC026312.1	GI:20070932			
VERSION	BC026312.1				
KEYWORDS	HTC.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 4039)				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.				
TITLE	Strausberg, R.				
JOURNAL	Direct Submission				
Submitted (02-APR-2002)	National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: cgapps-r@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu				

THIS PAGE BLANK (USPTO)

PN DE20103510-UI.
XX
PD 07-JUN-2001.
XX
PF 28-FEB-2001; 2001DE-2003510.
XX
PR 02-DEC-1999; 99DE-1058160.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
DR WPI; 2001-368570/39.
XX
XX
PT Gene library containing sequences with specific 3'-ends and no polyA
PT tail, encoding proteins involved in a wide range of cellular processes
PT
XX
PS
XX
XX Claim 15; Page 174; 251pp; German.
CC This invention describes a novel gene library (A) comprises a gene
CC sequence (or its part) encoding a protein involved in amino acid
CC synthesis, cellular/energy metabolism, metabolism of
CC fatty acids/phospholipids, synthesis or breakdown of
CC purines/pyrimidines/nucleosides/nucleotides, DNA
CC replication/transcription/translation, or is a transport/binding protein.
CC (A) are produced that correspond to the 3'-end of mRNA but without the
CC polyA tail. They can be prepared more efficiently and with less effort
CC than conventional libraries. AAK53436-AAK54275 represent fragments of the
CC gene library described in the method of the invention.
XX
SQ Sequence 327 BP; 71 A; 65 C; 84 G; 107 T; 0 other;

Query Match 18.8%; Score 251.8; DB 22; Length 327;
Best Local Similarity 92.6%; Pred. No. 5.4e-57;
Matches 299; Conservative 0; Mismatches 17; Indels 7; Gaps 3;

QY 981 CCACACACTTGCTCCA--GCCCCCACCCTAGCGGTTGGTGCATGAATGC-TTTTCGGCC 1037
DB 323 CCACACACTTGCTCCAAGCCCCCACCCTAGCGGTTGGTGCATGAATGCTTTTGGCC 264
QY 1038 CTTGACCGATGCCAAGCTAGGAGCCATGCTGATGATGTTGGTCTCGTCTTTCTTCATT 1097
DB 263 CTTGACCGATGCCAAGCTAGGAGCCATGCTGATGATGTTGGTCTCGTCTTTCTTCATT 204
QY 1098 GCAGGCTAGTATATACCAAGGGGATACAAACCGTTTGTCTCCATATTAATGATAAAACC 1157
DB 203 GCAGGCTAGTATATACCGAGGGGATACAAACCGTTTGTCTCCATATTAATGATAAAACC 144
QY 1158 ATTAACAACAAGCAAAACAAACA-----AAAACAAACAAACCAACCTTCCAATGTG 1213
DB 143 ATTAACAACAACAACAATAAACAACAACAACAACAACAACAACAACAACAACAACA 84
QY 1214 GGGAGAGAGAGAGCTTTCCGAGGCCGAGTGTTCGACATGCAGTAGGACATCATT 1273
DB 83 GGGAGAGAGAGAGCTTTCCGAGGCCGAGTGTTCGACATGCAGTAGGACATCATGTT 24
QY 1274 AGCAACTCAAGAAACAAACGAAA 1296
DB 23 AGCAACTCAAGAAACAAACGAAA 1

Search completed: March 15, 2003, 12:51:47
Job time : 329.202 secs

human RNA binding proteins which play a role in cancer, immune disorders and developmental disorders -

Examples; Columns 51-52; 39pp; English.

The present invention provides human RNA binding proteins (RNABP) and polynucleotides encoding the proteins. The polypeptides are useful for modulating the expression of human RNA binding proteins (RNABP) which play a role in cancer, immune disorders and developmental disorders. Disorders associated with a decrease of RNABP include: cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; immune disorders such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hyperosinophilia, irritable bowel syndrome, multiple sclerosis, osteoarthritis, osteoporosis, pancreaticitis, polynovitis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, X-linked agammaglobulinemia of Bruton, common variable immunodeficiency (CVI), Digeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated immunoglobulin (Ig)-A deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; and developmental disorders such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, sensorineural hearing loss, and any disorder associated with cell growth and differentiation, embryogenesis, and morphogenesis involving any tissue, organ, or system of a subject, e.g., the brain, adrenal gland, kidney, skeletal or reproductive system. The present sequence represents a cDNA encoding the RNABP-2 polypeptide.

Sequence 1506 BP; 452 A; 362 C; 377 G; 315 T; 0 other;

Query Match 20.6%; Score 275.6; DB 21; Length 1506;
Best Local Similarity 60.2%; Pred No. 5e-63;
Matches 601; Conservative 0; Mismatches 354; Indels 43; Gaps 7;

QY 222 GACAGATGATGCGCCCGACGCGGCGAGC---CCCAGACACACCTTCGAAAAACAC 278
DB 452 GACAGAGGTGGAGCAGACAGACGCGGCGAGCAGTCACAGACACAAAGTAGTGAATTC 511
QY 279 AGAAACAAGTCCAGCCCAAGCGGTGATGTCTCAACATCCCTTCGCGTTCGCGGA 338
DB 512 AGAGAGTAAATCTACCCGAAACGGCTGATGTCTCTAATATTCCTTCGCGTTCGCGGA 571
QY 339 TCCAGACCTCCGCAAAATGTTGGCCAAATTTGGTAAATATTAGATGTTGAAATATTATTT 398
DB 572 CCTTGACCTCCGCGAGATGTTGGGCGAGTTGGCAAAATCTTAGATGTAGAAATATCTT 631
QY 399 TAATGAGCGGGCTCGAAGGGATTTGGTTTCGTAACCTTTGAAAAATAGTGGCGATCGGA 458

Db 632 TAATGAACCTGGCTTAAGGGATTTCGGTTTCGTAACCTTTTCGAGATAGTCTGATGCAGA 691
QY 459 CAGGCGCAGGAGAAATTCACAGGTACCGTGTGTAGAGGCCGTAAATCGAGTTTAATAA 518
DB 692 CAGGCGCAGGAGAAATTCACAGGTACCGTGTGTAGAGGCCGTAAATCGAGTTGAATAA 751
QY 519 TGGCAGCAGCAGCGGTGATGACTTAATAAAGGCCGTGAACCCCTACACCAATGGCTGGAA 578
DB 752 TGTACAGCAGCGTGTATGACCAATGAAGATGGTGCACCAATATGCAAAATGTTGGAA 811
QY 579 ATTAATCCAGTTTGGCGCGGTCTACAGCCCGGACTTCTATGCAG-----GCA 628
DB 812 ATTAAGCCCAAGTAGTTGGAGCTGTATATGTCGGAGTTATATGACAGCATCCAGCTTCA 871
QY 629 CGGTGCTGTTTGTGCCAGGCCAACAGAGGGATTCCTTCATGTA--CAGTGGCCCCAGTTC 686
DB 872 AGCAGATGTGTCCTAGGCAATGATCAGCAGTGCCTTATCAGGAAGAGGGGTATCAA 931
QY 687 ACTTGATATATCTTCTGCAATGCTGCTTTCATA---TCGCGCGCCACTCTGTCAGC 743
DB 932 CACTTACATTCCTTAAATCAATTCCTGCTTCCCTTACCCTACTGACAGCCACACGGGAGC 991
QY 744 TGCATACCGAGGSGCTCACCTTCGAGGCGGTGTCGACCGCTGTACAAACCTTTCAGAGC 803
DB 992 CGCTTTCAGAGGAGCCCATTTGAGGGGAGAGGGCGGACAGTATATGTCGAGTCCGA-- 1049
QY 804 TGGCGGCGCCCGCCACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCCAGTGT 863
DB 1050 -GCGGTACTTCAACAGCCATCCCGCTATCCAGGGGTGATATGAGCTACAG---- 1104
QY 864 TGGCAATAAATTCGTACAGGCTGTTAGCTGTATACCGCTACGCCCGCCACCCCTGTC 923
DB 1105 -----ATATGCACAGCTGCTACTGCAACCGCAGCCACCGTGTCTGTCAGC 1149
QY 924 CACTGTGCTGCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 983
DB 1150 CGCTGACGCGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1209
QY 984 CACACTTGTCTCAGCCCCCAGCTACGCGTGTGGTGCATGAATGCTTTTGGGCCCTTGAC 1043
DB 1210 C---CTTGGCCCTGCGCTAGCTATGAGTGTGGCTGTGGCGAGTTTATATACCGAGTGG 1266
QY 1044 CGATGCCAAGCTAGGAGCCATGCTGATGATGTGGTCTGTTCTTCTTCTTCTTCTTCTT 1103
DB 1267 CTACAGCCGATTTTCCCGCTTACTGAAGTACGTCAGAGCCCTGCAAAATGGGACAGCCCC 1326
QY 1104 TAGTATATACCAAGGGGATACACCCGTTTGTCTCATATTTAAATGATAAAACCAATTA 1163
DB 1327 CAGTTCTAGGCGCTGCTGTTGCAATATTTACTAGTAGAGGAATCTTATAGCAAGATGA 1386
QY 1164 CAAACAAGCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1201
DB 1387 AGAGGAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1424

RESULT 15

AAK54001/c

ID AAK54001 standard; cDNA; 327 BP.

XX AAK54001;

AC AAK54001;

XX AAK54001;

DT 16-NOV-2001 (first entry)

XX Murine transcription associated protein encoding cDNA SEQ ID 566.

DE Murine; liver; gene library; amino acid synthesis; binding protein;

XX cell metabolism; energy metabolism; fatty acid metabolism; synthesis;

KW phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;

KW replication; transcription; translation; transport protein; ss.

XX Mus musculus.

OS Mus musculus.

XX Mus musculus.

PF 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PSDB; AAM40092.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 1451; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX
XX Sequence 1011 BP; 260 A; 314 C; 267 G; 170 T; 0 other;
SQ

Query Match 21.4%; Score 286.6; DB 22; Length 1011;
Best Local Similarity 64.6%; Pred. No. 5e-66;
Matches 574; Conservative 0; Mismatches 219; Indels 96; Gaps 6;
QY 155 CCTACACAGACGACGCGGAGCAGAGCGCTTATACAGCGCACAGCGCTCCGGCACCG 214
DB 157 CCCACCCCGACGACGCGGCTCCGAGGCCAGCACAGAGCCCATCGCGGGACCCAGACAG 216
QY 215 CCACACAGACAGATGATCGCCGCCACGACGCGGCGGCGGACACACACCTTCGAA 274
DB 217 TGCCGCGACAGACAGAGCGGCGGACAGCGGACGACGCGCGCT---CCACCCCTCCGACC 273
QY 275 ACACAGAAACAGTCCCGACGCGGCGGCTGATGTGTCACACATCCCTCCGCGTCC 334
DB 274 CTACAGAGAAGCAG---CAGCCCAAGCGGCTACACGTCTCCACATCCCTCCGCGTTCA 330
QY 335 GGGATCCAGACCTCCGACAAATGTTTGGCCAAATTTGGTAAATATATAGATGTGAAATTA 394
DB 331 GGGACCCCGACTTGGCGCAATGTTCCGGCAATTCGGAATAATTTAGACGTGGAGATCA 390
QY 395 TTTTAAATGAGCGGGCTCGAAGGATTTGGTTTCGTTACTTTCGAAATAGTGGGATG 454
DB 391 TTTTAAACGAGCGGGCTCCAGGGTTTGGGTTTGTAACTTTTGAACACTAGCTCAGATG 450
QY 455 CGGACAGGCGGAGGAGAAATTCACCGTACCGTGTAGAGGCGCGTAAATTCGAGGTTA 514
DB 451 CTGACCGAGCCCGGAGAGAGCTGATGGGACGATCGTAGAGGAGCGGAAATTTAGGTCA 510
QY 515 ATAATCGACAGACGCGGTGATGACTAATAAAGAGCCGCGTGAACCCCTACACCAATGGCT 574

DB 511 ATAATGCCAGCGCCCGAGTGATGACCAACAAGAACGCGGAACCCCTACCAACGGCT 570
QY 575 GGAATTAATTAATCAGTTGGGCGGCTCTACAGCCCCGACTTCTATGAGGACACGGTGC 634
DB 571 GGAAGCTAAATCAGTGGTGGCGCAGCTACGGGCTGAAATTTCTATGCAG----- 621
QY 635 TGTGTGCCAGGCGCAACACGAGGAGGATCTTCCATGTACAGTGGCCCCCAGTTCACTTGTAT 694
DB 622 ----- 621
QY 695 ATACTTCTGCAATGCTGCTTTCCATATCCGCGCCCACTGCTGCAGTGCATACCGAG 754
DB 622 -----TGACGGGGTTCCCTTACCCACACCGGACGCGGTTGCTTACCGGG 669
QY 755 GGGCTCACCTTCGAGGCGGCTGTCGCACCGTGTACAAACCTTCAGAGCTGCGGCGCCCC 814
DB 670 GCGCACATCTTCGGGGCCGGGGCGGTGTATATATACATTTTCGGGCTGCGCCACCCC 729
QY 815 CACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATAAT 874
DB 730 CACCCCAATCCCGGCTATGCGGAGCGGTGCTGTATCAGATGGATTTTATGTGCTGA-- 787
QY 875 TGCTACAGGCTGTTACGCTGCTACCGCTACCGCCAGCCCACTGCTGCTGCTGCTG 934
DB 788 -GATTTATGGAGGCTACGCGCTACAGATACGCTCAGC---CCGCTGCGAGCGGGCGCAG 843
QY 935 CCTACAGTACAGATTACGCGAGCTTTA---TGCTGCGGACCCCTACCAACACACTTGG 991
DB 844 CTTACAGCGACAGTATTACGCGAGCTTACGAGCTGCGGACCCGTACCATCACACCATCG 903
QY 992 CTCCAGCCCCACCTACGCGGTTGGTGCCATCAATGCTTTTCGCGCCTT 1040
DB 904 GCGCGCGGCGACCTACAGCATTTGGAACCATGTGAAACCTTCCACCGTT 952
RESULT 14
AAZ46827
ID AAZ46827 standard; cDNA; 1506 BP.
XX
XX AC AAZ46827;
XX
XX DT 11-APR-2000 (first entry)
XX
XX DE Human RNA binding protein (RNABP)-2 encoding cDNA (clone 1250374).
XX
XX KW RNA binding protein; RNABP; cancer; immune disorder; AIDS; human;
KW developmental disorder; acquired immunodeficiency syndrome; RNABP-2;
KW inflammation; allergy; diabetes mellitus; seizure disorder; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
FT CDS 201..1301
FT /*tag= a
FT /product= "RNABP-2"
XX
XX PN US6020164-A.
XX
XX PD 01-FEB-2000.
XX
XX PF 21-OCT-1998; 98US-0176657.
XX
XX PR 21-OCT-1998; 98US-0176657.
XX
XX PA (INCY-) INCYTE PHARM INC.
XX
XX PI Bandman O, Corley NC, Lu DAM, Baughn MR, Tang YT, Guegler KJ;
XX
XX DR WPI: 2000-146885/13.
XX
XX DR P-PSDB; AAY56850.
XX
XX Isolated and purified polynucleotide for modulating the expression of

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Auandi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; NAM41879.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 5024; 10078pp; English.
 XX
 SS The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (NAM38642-NAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;
 XX

Query Match 21.4%; Score 286.6; DB 22; Length 939;
 Best Local Similarity 64.6%; Pred. No. 4.8e-66;
 Matches 574; Conservative 0; Mismatches 219; Indels 96; Gaps 6;
 XX

QY 155 CTACACAGCAGCCTCGGAGCAGAGCGCTTATACAAAGCGGCACAGCGCTCTCCGGCAGC 214
 DB 875 CCACCCCGAGCGCAGGCTCGAGGCCAGCAGCAGCGCCATCGCGGGAGCCAGACAG 816
 QY 215 CCACACAGCAGATGATGCGGCCCGCCAGCGCCAGCCCGCCAGACACAACTTCTGAAA 274
 DB 815 TGCCGCGAGCAGCAGCGCGGCACAGCAGCAGCGCCGCT---CCACCCCTCCGACC 759
 QY 275 ACACAGAAACAGTCCCGAGCCCGCGCTGATGTGTCACATCCCTTCCGTTCC 334
 DB 758 CTACAGAGAAGCAG---CAGCCCAAGCGGTACACGCTCTCAACATCCCTTCCCGTTCA 702

QY 335 GGATCCAGACCTCCGACAAATGTTTGGCCAATTTGGTAAATATATAGATGTTGAATTA 394
 DB 701 GGGACCCCGAATTTCGGGCAAAATGTTTCGGCAATTCGGAATAATTTTAGCGTGGAGATCA 642
 QY 395 TTTTAAATGAGCGGGGCTCGAAGGGATTTGGTTTCTGTAACCTTTTGGAAAAATAGTGGCGATG 454
 DB 641 TTTTAAACGAGCGGGCTTCAAGGGTTTGGGTTTGTAACTTTTGAACCTAGCTCAGATG 582
 QY 455 CGGACAGGGCGAGGAGAAATTTGCACGGTACCGTGTAGAGGCCGTAAATCGAGGTTA 514
 DB 581 CTGACCGAGCCCGGAGAGAGCTGAATGGGACGATCGTAGAGGAGCGAAATTTGAGGTCA 522
 QY 515 ATAATGGACAGCAGCGGTGATGACTAATAAAGAGCGGTGAACCCCTACACCAATGGCT 574
 DB 521 ATAATGCCACGCGCCGAGTGATGACCAAGAAGACGGGGAACCCCTACACCAACGGCT 462
 QY 575 GGAATTAATCCAGTTGTGGGCGGTCTACAGCCCCGACTTCTATGACAGGCACGGTGC 634
 DB 461 GGAAGCTAATTCAGTTGTGGGCGAGTCTACGGGCTGNAATTCATGACG----- 411
 QY 635 TGTGTGCCAGGCCAACACAGGAGGATCTTCCATGTACAGTGGCCCCAGTTCACTTGTAT 694
 DB 410 ----- 411
 QY 695 ATACTTCTGCAATGCTGGCTTTCCATATCCGGCGCCACTGCTGCAGCTGCATACCGAG 754
 DB 410 -----TGACGGGGTTCCCTTACCCACACCGGCACAGCGGTTGCTTACCGGG 363
 QY 755 GGGCTACCTTTCAGAGCGGTGTCGACACCGGTACAAACACTTTCAGAGCTGCGGCGCCCC 814
 DB 362 GCGCACATCTTCGGGCGCGGCGCGGCGGTGTATATACATTTTCGGGCTGCGCCACCCC 303
 QY 815 CACCCCAATCCCGGCTATGGCGAGTAGTGTATCAAGAGCCAGTGTATGGCAATAAAT 874
 DB 302 CACCCCAATCCCGGCTATGGCGAGCGGTGCTGTATCAGGATGATTTTATGCTGCTGA-- 245
 QY 875 TGCTACAGGTTGTTACGCTGCATACCGCTACCGCAGCCACCCCTGCGACCTGCTGCTG 934
 DB 244 -GATTTATGAGGCTACGACCTACAGATACGCTCAGC---CGCTGCGAGCGCGGCGAG 189
 QY 935 CCTACAGTGACAGTTACGGACGAGTTTA---TGCTGCCGACCCCTACACCAACACACTTG 991
 DB 188 CCTACAGGACAGTTACGGACGAGTCTACGAGCTGCGGACCCGCTACCATCAGCAGTTCG 129
 QY 992 CTCAGGCCCCCAGCTACGCGGTTGTTGCCATGAATGCTTTTGGGCCCTT 1040
 DB 128 GGCCCGCGGCGACCTACAGCATTGGAACCATGTGAACCTTCCACCCGTT 80

RESULT 13
 AAI59248
 ID AAI59248 standard; cDNA; 1011 BP.
 XX
 AC AAI59248;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1451.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200153312-A1.
 XX
 PD 26-JUL-2001.
 XX

Qy 739 GCAGCTGCATACCGAGGGCTACCTTCGAGCGCGTGTGCGACCGGTGATCAACACCTTC 798
Db 1 GCGGCGGCTACCGAGGGCGCACCTGCGAGCGCGGTGCGACCGGTGATCAACACCTTC 60
Qy 799 AGACTGGGCGGCGCCACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCCA 858
Db 61 AGGCGCGGCGGCGCGCGCCCGCATCCCGGCTACGCGGTGTGTATCCAGGATGGA 120
Qy 859 GTGTATGCAATAAATTGCTACAGGCTGGTTACGCTGCATACCGCTAGCGCCAGCCACCC 918
Db 121 TTTTATGG--TGCAGACATTTATGGTGTATGCTGCATACCGCTAGCGCCAGCTTACC 177
Qy 919 CCTGCCACTGTGCTGCTCAGTGACAGTTACGACAGAGTTTATG--CTGCCAGCCCT 976
Db 178 CTGCGCACTGCGGCTGCTCAGTGACAGTTACGACAGAGTTTATGACATTTTCCCGCT 237
Qy 977 ACCACCACTGCTCCAGCGCCCACTACGCGGTGGTGGCCATGAATCTTTTGGC 1036
Db 238 GTCACCACTCAGTGTCTCAGCGCCCACTACGCGGTGGTGGCCATGAATCTTTTGGC 297
Qy 1037 CTTTGACCGATGCCAAGACTAGGAGCCATGCTGTATGATGTGGGTCTCTCTTTCTTCAT 1096
Db 298 CTTTGACTGATGCCAAGACTAGGAGCCATGCTGTATGATGTGGGTCTCTCTTTCTTCAT 357
Qy 1097 TGCAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCATATTAATGATAAACC 1156
Db 358 TGCAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCATATTAATGATAAACC 417
Qy 1157 CATTAACAAACAGCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1216
Db 418 CATAAA-----AACCTTCCAATGTGGG 440
Qy 1217 AGAGAGGAAGCTTTCGAGGCGCGAGTGTGCGACACATGC 1257
Db 441 AGAAGGAAGCTTTCGAGGCGCGAGTGTGCGACATGC 481
RESULT 10
ABK34530/C
ID ABK34530 standard; cDNA; 539 BP.
XX
AC ABK34530;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA for novel secreted protein, SEQ ID 299.
XX
KW Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
PN W020017290-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10295.
XX
PR 06-APR-2000; 2000US-194941P.
XX
PA (GENY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fichtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
XX WPI; 2002-179323/23.
XX

PT Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT
XX
PS Claim 1; Page 153; 339pp; English.
XX
CC The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins, their complements and sequences that hybridise to them.
CC Also included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the
CC polynucleotides, antibodies that bind to the proteins and identification
CC of modulators of the proteins or the expression of the polynucleotide.
CC The polynucleotides can be used as probes for the identification
CC and isolation of full length cDNA and genomic DNA. The polynucleotides
CC and proteins can also be used as nutritional supplements. The protein
CC is useful in the treatment of various immune deficiencies and disorders
CC such as viral infections, bacterial infections, fungal infections,
CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
CC autoimmune thyroiditis and diabetes) and allergic reactions and
CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.
XX
SQ Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;
Query Match 21.6%; Score 290; DB 24; Length 539;
Best Local Similarity 76.1%; Pred. No. 4.6e-67;
Matches 440; Conservative 0; Mismatches 45; Indels 93; Gaps 3;
Qy 748 TACCGAGGGGCTACCTTCGAGGCGGTGTCGACCGTGTACAAACCTTCAGAGCTGCG 807
Db 538 TACCGAGGGGCGCACCTTCGAGGCGCGGTGTCGACCGTGTACAAACCTTCAGGCGCGC 479
Qy 808 GCGCCCGCCACCCCAATCCCGGCTATGGCGGAGTAGTGTATCAAGAGCCAGTGTATGGC 867
Db 478 GCGCCCGCGCGCGCGCGCGCTACGCGGTGTGTTTACAGGATGGATTTTATGG- 420
Qy 868 AATAAATTGCTACAGGGTGTTCAGCTGCATACCGCTACGCCAGCCACCCCTGCCACT 927
Db 419 --TGCAGACATTTATGTTGTTATGCTGCATACCGCTACGCCAGCTTACCCCTGCCACT 362
Qy 928 GCTGCTGCTTACAGTGA-----TTCGTTCTTCGTTGCGAGAGATGAATTTCTTGT 944
Db 361 GCCGCTGCCTACAGTGACAGAAATCAGTTCTTCGTTGCGAGAGATGAATTTCTTGT 302
Qy 945 -----CAGTTACGAGAGAGTTTATGCTGCGCGAGCTTACCCACACACCTTGTCTC 994
Db 301 AACACCTCTGCGATGATGCGGCTTATGCTGCGAGAGTGTATGCTGCGAGCTTGTGCTG 242
Qy 995 CAGCCCGCCACCTACGCGGTTGGTGCCATGAATGCTTTTGGCCCTTGACCGATCCCAAGA 1054
Db 241 CAGCCCGCCACCTACGCGGTTGGTGCCATGAATGCTTTTGGCCCTTGACCTGATGCCAAGA 182
Qy 1055 CTAGGAGCCATGCTGATGATGCGGTCTCGTTCTTCTTCAATGCGAGCTAGTATATACC 1114
Db 181 CTAGGAGCCATGCTGATGATGCGGTCTCGTTCTTCTTCAATGCGAGCTAGTATATACC 122
Qy 1115 AAGGGGATACACCGTTTCTCCATATTAATGATAAACCATTAAACAAACAAAGCA 1174
Db 121 GAGGGGATACACCGTTTCTCCATATTAATGATAAACCATTAAACAAACAAAGCA 74
Qy 1175 AAAACAAACAAACAAACAAACAAACAAACCTTCCAAATGTTGGGAGAGAGAGCTTCGGA 1234
Db 73 -----NACCTTCCAATGTTGGGAGAGAGAGCTTCGGA 39

Db 842 CAGACCAAGGTTCCCGACACACATTAACCTGACCTCCCGCCCAACGACCTCCGAG 783
Qy 176 CAGAGC---GCTTATACAAAGCCACAG-CCGTCTCCGACCCGACACAGACAGATGAT 231
Db 782 CAGAGCCCGGCGACACAGCGCTCAGACCGTCTCTGACCCGACACACAGATAC 723
Qy 232 GCCGCCCCGACCGACCGCCACGCCGACACACACTTCTGAAAACACAGAAAACAGTCC 291
Db 722 GCACGACCGACGATGGCAGCGACCCGACACACACTTCTGAAAACACAGAAAACAGTCT 663
Qy 292 CAGCCCAAGCGCTGACATGTGTCCAACTCCCTTCGCGTCCGCGATCCAGACTCCGA 351
Db 662 CAGCCCAAGCGCTGACATGTGTCCAACTCCCTTCGCGTCCGCGATCCAGACTCCGA 603
Qy 352 CAAATGTTGSCCAATTTGTAATAATTAAGATGTTAAATTTTAAATGAGCGGCGC 411
Db 602 CAAATGTTGTCATTTGTAATAATTTGTAATAATTTTAAATGAGCGGCGC 543
Qy 412 TCGAAGGATTTGTTGTTCTGTAACCTTTCGAAAATAGTGGGATGCGGACGCGAGAG 471
Db 542 TCGAAGGATTTGTTGTTCTGTAACCTTTCGAAAATAGTGGGATGCGGACGCGAGAG 483
Qy 472 AAATTCACGCTACCGTGTAGAGGCGCTGTAATAATGAGGTTAAATGCGACAGCAGC 531
Db 482 AAATTCACGCGACCGTGTAGAGGCGCTGTAATAATGAGGTTAAATGCGACAGCAGC 423
Qy 532 GTGATGACTAATAAAAGCGCTGAAACCCCTACACCAATGCG 573
Db 422 GTATGACAAATTAATAAGACCGCGGCTCTGCTGGAGTGC 381

RESULT 6

AAKS3229/c
ID AAK53229 standard; cDNA; 1164 BP.

AAKS3229;

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 2758.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
tissue growth factor; immunomodulatory; cancer; leukemia;
nervous system disorder; arthritis; inflammation; ss.

Homo sapiens.

MO200157190-A2.

09-AUG-2001.

05-FEB-2001; 2001MO-US04098.

03-FEB-2000; 2000US-0496914.

27-APR-2000; 2000US-0560875.

20-JUN-2000; 2000US-0598075.

19-JUL-2000; 2000US-0620325.

01-SEP-2000; 2000US-0654936.

15-SEP-2000; 2000US-0663561.

20-OCT-2000; 2000US-0693325.

30-NOV-2000; 2000US-0728422.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW,
Xue AJ, Yang Y, Wejhrman T, Goodrich R,
WPI, 2001-476283/51.
P-PSDB; AAM80096.

PT Nucleic acids encoding polypeptides with cytokine-like activities.
PT useful in diagnosis and gene therapy -
PS Claim 1; Page 4968; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78333-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation and which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;

Query Match 27.5%; Score 368.4; DB 22; Length 1164;

Best Local Similarity 84.9%; Pred. No. 9e-88;

Matches 601; Conservative 0; Mismatches 86; Indels 21; Gaps 16;

Qy 1 GCGGCGCCGCTGACACAAATGCTCAG-CTTATGCTCAGCGAGTTCGCTCCACCCAG 59
Db 827 GCGGCTGCTGACACAAATGCTCAGCTTACGCTTCGCGCCAGTTTGTCTCCCGCAG 768
Qy 60 AATGCAATCCT-GCAGAAACAGGCGCCCTCATCCCATCCCGCCAGAG-ACACCGGC 117
Db 767 AACGATATCCCGCGGAATACAGCGCCCTCATCCCATCCCGCCAGAGTACACAGGC 708
Qy 118 CAGCCCATGTCCTCCCG-CCACACATTAATTTG-ATGCTCTTACAGAGGACTCGGAG 175
Db 707 CAGACAGGTTCCCGACACATTAATTTGATCTTCCCGCCAGAGCACTCCGAG 648
Qy 176 CAGAGC---GCTTATACAAAGCCACAG-CCGTCTCCGCGA-CCGCGACACAGACAGTGA 230
Db 647 CAGAGCCCGGCGACACAGACGCTCAGACCGTCTCTGCGACCCGCAACAAAGATTGA 588
Qy 231 TGCCGCCCCGACCGACCGCCACGACAC---AACTTCTGAAAACACAGAAAACAA 287
Db 587 CGAGCACCCACGAGTGGCCCAAGCCCAAGACAAACTTCTGAAAACACAGAAAACAA 528
Qy 288 GTC-CCAGCCCAAGC-GCTGCAATGTGTCCAATCCCTTC-CCGTTCCGGGATCCAGA 344
Db 527 GTCTCCAGCCCAAGCGGCTGCAATGTCTCCAATATCCCTTCAGGGTCCGGATCCGGA 468
Qy 345 CCTCCGACAAATG-TTGGCCCAATTTGTAATAATTTAGATGTTGAATTTTATATG 403
Db 467 CTTGACAAATGTTTGTGCAATTTGTAATAATTTAGATGTTGAATTTTATATG 408
Qy 404 AGCGGGCTCGAAGGATTTGTTGTAACCTTTCGAAAATAGTGGATGCGACAGCG 463
Db 407 AGCGAGGCTCAAGGATTTGTTGTAACCTTTCGAAAATAGTGGATGCGACAGCG 348
Qy 464 CGAGGAGAAA-TTGCACGTTACCGTGG-TAGAGGCGCGTAAATGAGTTAA-TAAT 519
Db 347 CGAGGAGAAATTTACCGGACCGTGTGTAAGAGGCGGTAATAATTTGAGTTAAAT 288
Qy 520 GCGACAGACGCGTATGATTAATAAAGCGGTGAAACCCCTACACCAATGCTGAGAA 579
Db 287 GCGACAGACGCTGATGATTAATAAAGACGCTACACCTTATACAAATGCTGAGAA 228
Qy 580 TTAATCAATGTTGGCGCGGCTACAGCCCGACTTCTATGACAGGACGCTGTTG 639
Db 227 TTGAATCAATGTTGGTGTGAGTCTACAGTCCGAATTTATGACAGGACGCTCTGTT 168
Qy 640 TGCCAGGCAACAGAGGATTTCCATGTACAGTGGCCCGCAGTTCA 687
Db 167 TGCAGGCAACAGAGGATTTCCATGTACAGTGGCCCGCAGCA 120

```

Db 1927 GCTGCATACCGCTACGCCAGCTTACCCCTGCGACTGCGCTGCTACAGTACAGAAAT 1986
Qy 945 -----CAGTTACGGACGAG 958
Db 1987 CAGTTGCTTCTGCTGACGACATGAAATTTCTTGAACACTCTGCGAGTTACGACAG 2046
Qy 959 TTTATGCTGCGGACCCCTACACACACACACTTGTCTCCAGCCCCCCTACGGCGTTGGTG 1018
Db 2047 TTTATGCTGCGGACCCCTACACACACACACTTGTCTCCAGCCCCCCTACGGCGTTGGTG 2106
Qy 1019 CCATGAATGCTTTTGGCGCTTGACCCGATGCCAAGACTAGAGAGCCATGCTGATGATGG 1078
Db 2107 CCATGAATGCTTTTGGCGCTTGACCCGATGCCAAGACTAGAGAGCCATGCTGATGATGG 2166
Qy 1079 GCTGCTCTTCTTCTCATTTGACGAGTATATATACCAAGGGGAGATACACGCTTTGCTC 1138
Db 2167 GCTGCTCTTCTTCTCATTTGACGAGTATATATACCAAGGGGAGATACACGCTTTGCTC 2226
Qy 1139 CATATTAAATGATTAACCATTTAAACAAACAGCAAAACAAACAAACAAACAAAC 1198
Db 2227 CATATTAAATGATTAACCATTTAAACAAACAGCAAAACAAACAAACAAACAAAC 2250
Qy 1199 CAACCTTCCATGTCGGGAGAGAGAGAGAGCTTCCGAGGCCGAGTGTTCGACACATGCA 1258
Db 2251 -AACCTTCCATGTCGGGAGAGAGAGAGAGCTTCCGAGGCCGAGTGTTCGACATGCA 2309
Qy 1259 GTAGGACATCACTTAGCACTCAAGAAACAAACGAAAAAATATAAATAAATAAATAA 1318
Db 2310 GTAGGACATCACTTAGCACTTAAAGCAATTTTAAAAAATAAATAAATAAATAAATAA 2369

```

RESULT 5

AAS31173/c
ID AAS31173 standard; cDNA; 1295 BP.

AC AAS31173;

DT 04-DEC-2001 (first entry)

DE Human diagnostic and therapeutic polynucleotide (DITHP) #188.

KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
acquired immune deficiency syndrome; AIDS; autoimmune disorder;

KW respiratory disorder; ss.

OS Homo sapiens.

PN WO200162927-A2.

PD 30-AUG-2001.

PF 21-FEB-2001; 2001WO-US06059.

XX 24-FEB-2000; 2000US-0184693.

PR 24-FEB-2000; 2000US-0184697.

PR 24-FEB-2000; 2000US-0184698.

PR 24-FEB-2000; 2000US-0184768.

PR 24-FEB-2000; 2000US-0184769.

PR 24-FEB-2000; 2000US-0184770.

PR 24-FEB-2000; 2000US-0184771.

PR 24-FEB-2000; 2000US-0184772.

PR 24-FEB-2000; 2000US-0184773.

PR 24-FEB-2000; 2000US-0184774.

PR 24-FEB-2000; 2000US-0184776.

PR 24-FEB-2000; 2000US-0184777.

PR 24-FEB-2000; 2000US-0184797.

PR 24-FEB-2000; 2000US-0184813.

PR 24-FEB-2000; 2000US-0184837.

PR 24-FEB-2000; 2000US-0184841.

PR 24-FEB-2000; 2000US-0185213.

PR 24-FEB-2000; 2000US-0185216.

PR 12-MAY-2000; 2000US-0203785.

```

PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204525.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC.
PI Chen A, D'Sa SA, Amshay S, Dahl CR, Dam TC, Daniels SE.
PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL.
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Datto A.
PI Wright RJ, Yap PE, Yu YJ, Bradley DL, Bratcher SR, Chen W.
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S.
XX
XX WPI, 2001-502867/55.
XX P-PSDB; AAU19602.
XX
XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
PT
XX
XX Claim 1; Page 385; 522pp; English.
XX
XX The invention relates to polynucleotides (I) encoding diagnostic and
XX therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
XX and proteins involved in growth and development and receptors. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate DITHP expression. For example, (I) and
XX (II) may be used to treat disorders associated with decreased polypeptide
XX expression by rectifying mutations or deletions in a patient's genome,
XX that affect the activity of the DITHPs, by expressing inactive proteins
XX or supplementing the patient's own production of them. (I) and (II)
XX may be used to treat diseases, for example, cell proliferative disorder,
XX Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
XX leukemia, autoimmune disorders, and respiratory disorders. Additionally,
XX (I) may be used to produce the DITHPs, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the protein. (I) and
XX its complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar nucleic acids in
XX samples, and therefore which patients may be in need of restorative
XX therapy. (II) may also be used as antigens in the production of DITHP
XX antibodies against DITHPs and in assays to identify modulators of DITHP
XX expression and activity. The anti-DITHP antibodies and antagonists may
XX also be used to down regulate expression and activity. The anti-DITHP
XX antibodies may also be used as diagnostic agents for detecting the
XX presence of DITHPs in samples (e.g. by enzyme linked immunosorbent
XX assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and
XX therapeutic (DITHP) polynucleotides of the invention.
XX
XX Sequence 1295 BP; 253 A; 372 C; 343 G; 326 T; 1 other.
SQ

```

Query Match 28.7%; Score 385.2; DB 22; Length 1295;

Best Local Similarity 86.8%; Pred. No. 3.2e-92;

Matches 505; Conservative 0; Mismatches 68; Indels 9; Gaps 7;

```

Qy 1 GGGGCGCGCCGACACATGCTGAG-CTTATGCTGACGCGAGTGGCTCCACCCGAG 59
Db 962 GCGGCTGCGCCGACACATGCTGAGCTTACGCTTGGAGCCAGTTGCTCCCGCAG 903
Qy 60 AATGCAATCCT-GCAGAAATACAGCGCCCTCATCCCATCCCGCCAGAG-ACACCGCG 117
Db 902 AACGATATCCCGCGGATATACAGCGCCCTCATCCCATCCCGCCAGAGATACAGAGCC 843
Qy 118 CAGCCCACTGTCCCG-CCACACATTAACCTTG-ATCTCTCTACACAGAGCATCTGGAG 175

```


DE Human cDNA sequence SEQ ID NO:10786.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
XX Claim 8; SEQ ID 10786; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95883 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;
XX
XX Query Match 67.6%; Score 905.2; DB 22; Length 1513;
XX Best Local Similarity 86.7%; Pred. No. 1.4e-230;
XX Matches 1118; Conservative 0; Mismatches 123; Indels 49; Gaps 9;
XX
Qy 1 GGGGCGGCGCCCTGACACATGCTCAG-CTTATGCTCAGCGCAGTTCCCTCACCCGAG 59
Db 263 GCGGCGGCGCCCTGACACATGCTCAGCTTACGCTTGGGCGGCGGAGTTCTCCCGCGAG 322
Qy 60 AATGCACTCT-GCAGATATACGCGCCCTCATCCCATCCGCGCGAGAG-ACACGCGC 117
Db 323 AACGGATATCCCGCGGAAATACAGCGCCCTCATCCCATCCCGCGCGAGAGTACAGCGC 382
Qy 118 GAGCCCACTGTCGCGG-CCACACATTAACCTTG-ATCTCTCTACACAGAGCGACTGCGAG 175
Db 383 CAGACCACTGTCGCGG-CCACACATTAACCTTGATCTCTCCGCGCGGAGAGCGACTGCGAG 442

Qy 176 CAGAGC---GCTTATACAGCGCACAG-CCGCTTCGCGCACCGCCACACAGACATGAT 231
Db 443 CAGAGCCCGCGGACAGAGCGCTCAGACCGCTCTGCGCACCGCCACAGAGATGAC 502
Qy 232 GCGGCGGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 291
Db 503 GAGACCGGAGATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 562
Qy 292 CAGCCCAAGCGGCTGATGTGTCCAGATCCCTTCGCGGTCGCGGATCCAGACTCCGA 351
Db 563 CAGCCCAAGCGGCTGATGTGTCCAGATCCCTTCGCGGTCGCGGATCCAGACTCCGA 622
Qy 352 CAATGTTGGCCCAATTTGTAATAATATAGATGTAATATTTTATATGAGCGGCGC 411
Db 623 CAATGTTGGCTCAATTTGTAATAATATAGATGTAATATTTTATATGAGCGAGCGC 682
Qy 412 TCGAAGGATTTGGTTGCTGTATCTTGAATAATGTGCGGATGGGAGCAGGCGGAGG 471
Db 683 TCGAAGGATTTGGTTGCTGTATCTTGAATAATGTGCGGATGGGAGCAGGCGGAGG 742
Qy 472 AAATTCACGCTACCGGTGTAGAGGCGGTAAATCGAGTTAATATGACAGACAGC 531
Db 743 AAATTCACGCGCACCGGTGTAGAGGCGGTAAATCGAGTTAATATGACAGACAGC 802
Qy 532 GTGATGACTAATTAAGGCGCGGTGAACCCCTACACCAATGCTGGAATTAATCACT 591
Db 803 GTAATGAATAATAAGGCGCGGTGAACCCCTACACCAATGCTGGAATTAATCACT 862
Qy 592 GTGGCGGCGGCTACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651
Db 863 GTGGCGGCGGCTACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 922
Qy 652 CAGAGGAGATTTCCATATAGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 711
Db 923 CAGAGGAGATTTCCATATAGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 982
Qy 712 GAGTTTCAATATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 771
Db 983 GAGTTTCCGTTTCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1042
Qy 772 CGTGTCGACCGGTATTAACACCTTCAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 831
Db 1043 CGGTCGACCGGTATTAACACCTTCAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1102
Qy 832 TATGCGGAGTATGATATCAAGAGCGAGTATGAGCAATTAATGCTACAGGCTGTTAC 891
Db 1103 TATGCGGAGTATGATATCAAGAGCGAGTATGAGCAATTAATGCTACAGGCTGTTAC 1159
Qy 892 GCTGATACCGGTATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 951
Db 1160 GCTGATACCGGTATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1219
Qy 952 GAGCAGATTTATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1011
Db 1220 GAGCAGATTTATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1279
Qy 1012 GTTGTGCGATGATGATGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1071
Db 1280 GTTGTGCGATGATGATGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1339
Qy 1340 GATGTGGGCTCTGCTTCTTCTTATGAGGCTATATATACCGGCGGCGGCGGCGGCGGCGG 1399
Qy 1132 TTTGCTCATATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1191
Db 1400 TTTGCTCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1430
Qy 1192 AAAAACAACCTTCAATGTGGGAGAGAGAAAGCTTTCGAGGCGCGGAGTGTGCGAC 1251
Db 1431 -----AACCTTCAATGTGGGAGAGAAAGACCTTTCGAGGCGCGGAGTGTGCGAAT 1482
Qy 1252 ACATGCAATGAGACATCACTTAGCACTC 1281

XX 09-AUG-2001.
 PD 05-FEB-2001; 2001WO-US04098.
 XX
 PF
 XX 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0649316.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Arundt V, Zhou P, Xu C, Gao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR P-PSDB; AAM79112.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 2643-2645; 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK51435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX
 SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
 Query Match 67.7%; Score 906.8; DB 22; Length 1800;
 Best Local Similarity 86.7%; Pred. No. 5.7e-231;
 Matches 1119; Conservative 0; Mismatches 122; Indels 49; Gaps 9;

QY 352 CAAATGTTGGCCCAATTGGTAAATATTAGATGTTAAATTTTAAATAGACGGGCGC 411
 DB 910 CAAATGTTGGTCAATTGGTAAATATTAGATGTTAAATTTTAAATAGACGGGCGC 969
 QY 412 TCGAAGGATTTGGTTCGTAATCTTGGAAATATAGTCGGATGCGACAGGAGGAG 471
 DB 970 TCAAGGATTTGGTTCGTAATCTTGGAAATATAGTCGGATGCGACAGGAGGAG 1029
 QY 472 AAATTGACGGTACCGTGTAGAGGCGGTAAATTCAGGTTAAATTCGACACGACGC 531
 DB 1030 AAATTACACGGACCGTGTAGAGGCGGTAAATTCAGGTTAAATTCGACACGACGC 1089
 QY 532 GTGATGCTAATTAATAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTT 591
 DB 1090 GTAAATGACAAATTAATAAGGCGGTGAACCCCTAATTAATGCTGGAATTAATTCAGTT 1149
 QY 592 GTGGGCGCGGTCTACACCGCCGACTTATAGACGGGCGGTGCTGTGCGACGGCCAC 651
 DB 1150 GTGGGTGACGTACAGATCCGAAATCTTATGACGGGCGGTGCTGTGCGACGGCCAC 1209
 QY 652 CAGAGGGATCTTCCATGTACAGTGGCCCGCAGTTCACTTGTATATACCTTGCATATGCT 711
 DB 1210 CAGAGGGATCTTCCATGTACAGTGGCCCGCAGTTCACTTGTATATACCTTGCATATGCT 1269
 QY 712 GGCCTTCCATATCCGCGCCGCACTGCTGACAGTGCATACGAGGGGCTCACCTTCAGGC 771
 DB 1270 GGCCTTCCGATATCAGACGACCGCCGCGCGCTTACGAGGGGCGCACCTTCGAGGC 1329
 QY 772 CGGTGTGACCGGTATCAACACCTTGAGAGTGGCGGCGCCCGCCCAATCCGGGCGC 831
 DB 1330 CGGTGTGACCGGTATCAACACCTTGAGAGTGGCGGCGCCCGCCCGGATCCGGGCGC 1389
 QY 832 TATGCGGAGTAGTATCAAGAGCCAGTATGCAATTAATTTCTACAGGGTGTAC 891
 DB 1390 TATGCGGAGTAGTATCAAGAGCCAGTATGCAATTAATTTCTACAGGGTGTAC 1446
 QY 892 GCTGATACCGCTACGCGCCAGCCACCTTGCACCTGCTGCTACAGTACAGTTAC 951
 DB 1447 GCTGATACCGCTACGCGCCAGCCACCTTGCACCTGCTGCTACAGTACAGTTAC 1506
 QY 952 GGAGGATTTATGTCGCGACCCCTTACACACATTTCTCCAGGCCCTACAGGC 1011
 DB 1507 GGAGGATTTATGTCGCGACCCCTTACACACATTTCTCCAGGCCCTACAGGC 1566
 QY 1012 GTTGTGTCATGAATGCTTTTGGCGCCCTTACAGGATCCCAAGACTAGAGCCATGCTAT 1071
 DB 1567 GTTGTGTCATGAATGCTTTTGGCGCCCTTACAGGATCCCAAGACTAGAGCCATGCTAT 1626
 QY 1072 GATGTGGGCTCGTCTTCTTCTTCAATTCAGGCTATATATACCAAGGGGATACACCGT 1131
 DB 1627 GATGTGGGCTCGTCTTCTTCTTCAATTCAGGCTATATATACCAAGGGGATACACCGT 1686
 QY 1132 TTTGCTCCATATTAATGAATTAACATTAACAAACAAACAAACAAACAAACAAACAA 1191
 DB 1687 TTTGCTCCATATTAATGAATTAACATTAACAAACAAACAAACAAACAAACAAACAA 1217
 QY 1192 AAAAAACCACTTCCATGTGGGAGAGAGAAAGCTTTCCGAGGCGCGAGTGTGGCAG 1251
 DB 1718 -----AACCTTCCAAATGTGGGAGAGAAAGCTTTCCGAGGCGCGAGTGTGGCAGT 1769
 QY 1252 ACATGAGTAGAGCATCATTTAGCAATC 1281
 DB 1770 ACATGAGTAGAGCATCATTTAGCAATC 1799

RESULT 3
 AAH13824
 ID AAH13824 standard; cDNA; 1513 BP.
 XX
 AC AAH13824;
 XX
 DT 26-JUN-2001 (first entry)
 XX

PA (SCIO-) SCIOS INC.
 XX Stanton LW, White RT;
 XX WPI, 2002-010779/01.
 DR P-PSDB; AAU070146.
 XX Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX Claim 1; Fig 1; 189pp; English.
 PS
 XX The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 XX
 SQ Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;
 Query Match 100.0%; Score 1340; DB 24; Length 1340;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 541 AATAAAAAGCCGTGAACCCCTACACCAATGGCTGGAATTAATTAATCCAGTTGTGGCCG 600
 Qy GTCTACAGCCCGCACTTCTATGACAGGACGGTGTGTGTGCCAGGCCAACAGAGGGA 660
 Db 601 GTCTACAGCCCGCACTTCTATGACAGGACGGTGTGTGTGCCAGGCCAACAGAGGGA 660
 Qy TCTTCATGTAAGTGGGCCCGAGTTCACTTGTATATACCTTGTGCAATGCTGGCTTTTCA 720
 Db 661 TCTTCATGTAAGTGGGCCCGAGTTCACTTGTATATACCTTGTGCAATGCTGGCTTTTCA 720
 Qy TATCCGGCCGCACTGCTGCACTGCAATACGAGGGGCTCACTTTCAGGCGGTGTGCG 780
 Db 721 TATCCGGCCGCACTGCTGCACTGCAATACGAGGGGCTCACTTTCAGGCGGTGTGCG 780
 Qy ACCGTGTCAACACCTTCAAGAGTGGCGGCCGCCACCCCAATCCCGCTATAGCCGA 840
 Db 781 ACCGTGTCAACACCTTCAAGAGTGGCGGCCGCCACCCCAATCCCGCTATAGCCGA 840
 Qy GTAGTGTATCAAGAGCCAGTGTATGGCAATTAATTTGCTACAGGGGTGTACGCTGCATAC 900
 Db 841 GTAGTGTATCAAGAGCCAGTGTATGGCAATTAATTTGCTACAGGGGTGTACGCTGCATAC 900
 Qy CGTACGCCGCAAGCCCAACCCCTGCACTGCTGCTGCTACAGTGAAGTACGAGCAAGTT 960
 Db 901 CGTACGCCGCAAGCCCAACCCCTGCACTGCTGCTGCTACAGTGAAGTACGAGCAAGTT 960
 Qy TATGCTGCCAGCCCTTACACCACTTGTCTCAGGCCCTACCTACGCGGCTGTGCTCC 1020
 Db 961 TATGCTGCCAGCCCTTACACCACTTGTCTCAGGCCCTACCTACGCGGCTGTGCTCC 1020
 Qy ATGAATGCTTTTGGCCCTTACCGATGCCAAGACTAGAGCCATGCTGATGATGTGGGT 1080
 Db 1021 ATGAATGCTTTTGGCCCTTACCGATGCCAAGACTAGAGCCATGCTGATGATGTGGGT 1080
 Qy CTGCTTTCTTTCTTCAATGCAAGCTAGTATATACCAAGGGGATCAACCGTTTGTCTCA 1140
 Db 1081 CTGCTTTCTTTCTTCAATGCAAGCTAGTATATACCAAGGGGATCAACCGTTTGTCTCA 1140
 Qy TATTAAATGATTAATAACATTAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1200
 Db 1141 TATTAAATGATTAATAACATTAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1200
 Qy ACCTTCAATGTGGGAGAGAGAAAGCTTTCCGAGGCCGAGTGTGGACACATGCAGT 1260
 Db 1201 ACCTTCAATGTGGGAGAGAGAAAGCTTTCCGAGGCCGAGTGTGGACACATGCAGT 1260
 Qy AGAATATCACTTTAGCACTCAAGAAACAAACGAAAAAATTAAGC 1320
 Db 1261 AGAATATCACTTTAGCACTCAAGAAACAAACGAAAAAATTAAGC 1320
 Qy GAGCGAAGGGGTGCTAGA 1340
 Db 1321 GAGCGAAGGGGTGCTAGA 1340

RESULT 2
 ID AAK52245 standard; cDNA; 1800 BP.
 XX AAK52245;
 AC 06-NOV-2001 (first entry)
 DT
 XX
 DE Human polynucleotide SEQ ID NO 790.
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 OS Homo sapiens.
 XX
 XX
 FN W0200157190-A2.

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:24:02 ; Search time 314.202 Seconds
(without alignments)
9604.255 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340
Sequence: 1 gcggcgcccccgcacacacat.....ggccgaagggttcgctaga 1340

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Genes2_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
25: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	1340	24	AAS94693
2	906.8	67.7	1800	22	AAK52245
3	905.2	67.6	1513	22	AAH13824
4	857.2	64.0	2372	21	AAH07075
5	385.2	28.7	1295	22	AAH31173
6	368.4	27.5	1164	22	AAH53129
7	338.8	25.3	607	22	AAH09205
8	334	24.9	893	22	AAH03678
9	319	23.8	481	23	AAH93633

C	10	290	21.6	539	24	ABK34530	Human CDNA for nov
C	11	286.6	21.4	939	22	AA161034	Human polynucleoti
C	12	286.6	21.4	939	22	AA161035	Human polynucleoti
C	13	286.6	21.4	1011	22	AA159248	Human polynucleoti
C	14	275.6	20.6	1506	21	AA246827	Human RNA binding
C	15	251.8	18.8	327	22	AAK54001	Murine transcripti
C	16	229	17.1	406	23	AAH53532	DNA encoding novel
C	17	170.6	12.7	3189	24	ABK34602	Human CDNA for nov
C	18	154	11.5	395	22	ABH11642	Human nervous syst
C	19	146	10.9	903	23	ABH13655	Drosophila melanog
C	20	145.4	10.6	394	20	AAH66245	DNA encoding novel
C	21	141.8	10.6	394	20	AAH87007	EST clone AY147.
C	22	134.4	10.0	918	22	AAH59249	Human polynucleoti
C	23	123.8	9.2	473	22	AAK07716	Human brain expres
C	24	123.8	9.2	473	22	AAK33549	Human bone marrow
C	25	123.8	9.2	473	22	AAH39304	Probe #17990 used t
C	26	123.8	9.2	473	24	ABH08416	Human genome-deriv
C	27	109.6	8.2	125	22	AAK20380	Human brain expres
C	28	109.6	8.2	125	22	AAK46486	Human bone marrow
C	29	109.6	8.2	125	22	AAH52354	Probe #21040 used
C	30	109.6	8.2	125	24	ABH20830	Human genome-deriv
C	31	107	8.0	2118	23	AAH93634	DNA encoding novel
C	32	93	6.9	1252	23	AAH90510	Human foetal liver
C	33	88	6.6	224	22	ABH70887	Probe #15799 for g
C	34	88	6.6	224	22	ABH37333	Human brain expres
C	35	88	6.6	224	22	AAK19155	Human bone marrow
C	36	88	6.6	224	22	AAK45114	Probe #15075 for g
C	37	88	6.6	224	22	AAH25142	Human genome-deriv
C	38	88	6.6	224	22	AAH51072	Probe #19758 used
C	39	88	6.6	224	24	ABH51934	Human genome-deriv
C	40	86.4	6.4	454	22	ABH58290	Human foetal liver
C	41	86.4	6.4	454	22	ABH27427	Probe #5893 for ge
C	42	86.4	6.4	454	22	AAK06386	Human brain expres
C	43	86.4	6.4	454	22	AAK32056	Human bone marrow
C	44	86.4	6.4	454	22	AAH15924	Probe #5857 for ge
C	45	86.4	6.4	454	22	AAH37912	Probe #6598 used t

ALIGNMENTS

RESULT 1	AAS94693	standard; cDNA; 1340 BP.
ID	AAS94693;	
XX	12-MAR-2002.	(first entry)
DT		
XX		
DE	Rat secreted factor DNA clone P0184_D11 #1.	
XX		
KW	Rat; secreted factor polypeptide; cardiac disease; renal disease; kidney;	
KW	inflammatory disease; congestive heart failure; myocarditis; asthma; ss;	
KW	dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;	
KW	myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;	
KW	atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;	
KW	renal infarction; hereditary nephritis; polycystic kidney disease;	
KW	chronic renal failure; renal vein thrombosis; medullary sponge kidney;	
KW	rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;	
KW	graft versus host reaction; Crohn's disease; ulcerative colitis; probe;	
KW	Alzheimer's disease; gene therapy.	
XX		
OS	Rattus norvegicus.	
XX		
PN	WO200174901-A2.	
XX		
PD	11-OCT-2001.	
XX		
XX	23-MAR-2001; 2001WO-US09555.	
PF		
XX		
PR	31-MAR-2000; 2000US-193548P.	
XX		
PR	14-MAR-2001; 2001US-0809545.	
XX		

Db 112435 TTAATGATTAACCATTAACAAACAAACAAACAAACAAACAAAC 112376
Qy 1203 CTTCCAATGTGGGAGAGAGAGAGCTTTCCGAGGCCCGAGTGTTCGACACATGCAGTAG 1262
Db 112375 CTTCCAATGTGGGAGAGAGAGAGCTTTCCGAGGCCCGAGTGTTCGACACATGCAGTAG 112316
Qy 1263 GACATCACTTTAGCACTCAAAAGAAACAAACGAAAAAATA 1317
Db 112315 GACATCACTTTAGCACTCAAAAGAAACAACTTAAAAAATCAAAAAA 112261

Search completed: March 15, 2003, 14:20:15
Job time : 3622.42 secs

REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 115860)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Briefe,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegod,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaikie,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanli,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 115860)
Worley,K.C.
Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115860)
Worley,K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20514522.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXLC
Center clone name: CH230-34C2
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 99693 bases at least Q40
Consensus quality: 101936 bases at least Q30
Consensus quality: 102743 bases at least Q20

TITLE
JOURNAL

Unpublished

2 (bases 1 to 115860)

Worley,K.C.

REFERENCE
AUTHORS

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115860)
Worley,K.C.

REFERENCE
AUTHORS

Direct Submission

Submitted (23-JUL-2002)

Human Genome Sequencing Center, Department

Of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 18, 2002 this sequence version replaced gi:20514522.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GXLC

Center clone name: CH230-34C2

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 99693 bases at least Q40

Consensus quality: 101936 bases at least Q30

Consensus quality: 102743 bases at least Q20

COMMENT

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 1347: contig of 1347 bp in length
1348 1447: gap of unknown length
1448 2834: contig of 1387 bp in length
2835 2935: gap of unknown length
2936 4253: contig of 1319 bp in length
4254 4354: gap of unknown length
4355 5477: contig of 1124 bp in length
5478 7484: gap of unknown length
7485 7584: contig of 1907 bp in length
7585 9406: gap of unknown length
9407 9507: contig of 1822 bp in length
9508 11712: contig of 2206 bp in length
11713 13462: gap of unknown length
13463 13562: contig of 1650 bp in length
13563 17222: contig of 3660 bp in length
17223 20618: gap of unknown length
20619 20717: contig of 3295 bp in length
20718 24825: contig of 4108 bp in length
24826 24925: gap of unknown length
24926 28812: contig of 3886 bp in length
28813 33686: gap of unknown length
33687 33786: contig of 4775 bp in length
33787 40217: contig of 6431 bp in length
40218 47912: gap of unknown length
47913 48012: gap of unknown length
48013 56482: contig of 8470 bp in length
56483 56582: gap of unknown length
56583 66204: contig of 9622 bp in length
66205 74546: gap of unknown length
74547 74646: contig of 8242 bp in length
74647 85222: gap of unknown length
85223 86623: contig of 11876 bp in length
86624 100049: contig of 13427 bp in length
100050 100149: gap of unknown length
100150 115860: contig of 15711 bp in length.

FEATURES
source

1. 115860
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-34C2"

BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others
ORIGIN

Query Match 21.3%; Score 285.4; DB 2; Length 115860;
Best Local Similarity 98.0%; Pred. No. 5.8e-66;
Matches 289; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1023 GAATGCTTTTGGCCCTTGACCGATGCCAGCTAGAGCCATGCTGATGTTGGTCT 1082
|||||
Db 112555 GAATGCTTTTGGCCCTTGACCGATGCCAGCTAGAGCCATGCTGATGTTGGTCT 112496
Qy 1083 CGTTCTTTTCTTTCATTGAGGCTAGTATATACCAAGGGGGATACAAACCGTTTGGTCCATA 1142
Db 112495 CGTTCTTTTCTTTCATTGAGGCTAGTATATACCAAGGGGGATACAAACCGTTTGGTCCATA 112436
Qy 1143 TTAATGATATAAACCATTAACAAACCAAGCAAAAAACAAACCAAAACCAAC 1202

```

QY 742 GCTGCATACCGAGGGCTCACCTTCGAGGCGGTGTCGACCGCTGTACACACCTTCAGA 801
Db 793 GCTGCTTCGAGAGAGCCATCTCAGGGGAGAGGGGCGAGCTGTATGTGTCAGTCCGA 852
QY 802 GCTGCGGCGCCGCCACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCCAGTG 861
Db 853 --GCGGTACCTCCAAACAGCATCCCGGCTATCCAGGTGTGGTTTACCAGGACGGATT 909
QY 862 TATGGCAATAATTGCTACAGGCTGTACGCTGATACGCTACGCTACGCTACGCTACGCT 921
Db 910 TACGGTCTGA---CCTCTATGTTGGATATGACGCTTACGAGCTTACGAGCTGTACT 966
QY 922 GCCA-----CTGCTGCTGCTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 957
Db 967 GCAACCGCAGCAGCTGCTGACGCGCTGACGCGCTTACAGCAGGTTACGGCAGG 1026
QY 958 GTTATGTCGCGACCCCTACACACACACTTGTCTCAGCCCCCCTACGCGGTGGT 1017
Db 1027 GTGTACACAGCTGACCCCTACCATGCC---CTCGCCCTGCCCGCAGCTATGGAGTTGGC 1083
QY 1018 GCCATG 1023
Db 1084 GCTGTG 1089

RESULT 14
AF229055 1695 bp mRNA linear ROD 01-FEB-2002
DEFINITION Mus musculus hexaribonucleotide binding protein 2 (Hrnpb2) mRNA,
complete cds.
ACCESSION AF229055
VERSION AF229055.1 GI:18461362
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1. (bases 1 to 1695)
AUTHORS Chen, W. and Winkelman, J.C.
TITLE Direct Submission
JOURNAL Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, The Vontz Center for
Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
Location/Qualifiers
FEATURES
source 1..1695
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..1695
/ gene="Hrnpb2"
200..1333
/ gene="Hrnpb2"
/ note="RNA-binding protein"
CDS /codon_start=1
/product="hexaribonucleotide binding protein 2"
/db_xref="GI:18461362"
/protein_id="AAL71902.1"
/translacion="MEKKNVTPGQNEPTTTPDMVQPTTIPFPFPFGIPEYGV
PHQYAGTSSHNLTLYGSTPHGEOSNSPNSGSLTOTEAGQTDGQOSQTS
ENSEKSTPKRLHVSNI PRFRDPLROMFGQFKLDVEIIFNERGSKGFGVTFEN
SADADRAREKLHGVVVEGRKIEVNNATRVMTNKMVTPYANGWKLSPVVGAVYGP
YAASFDVADVLSGNEAAVPMGSGINTYIPLIIPGFFPTATTAAPRGNHGRG
RTVYGAVRAVPTAIYPVGVYQDFYGGVYAAVRYAPATATATAAATAA
AYSDGVRVYTDPAHALAASVGVGAVASLYRGYSRFAPY"
BASE COUNT 450 a 395 c 432 g 418 t
ORIGIN
Query Match 22.0%; Score 295.2; DB 10; Length 1695;
Best Local Similarity 64.8%; Pred. No. 5.8e-69;
Matches 587; Conservative 0; Mismatches 268; Indels 51; Gaps 8;
QY 160 ACAGACCACTCGGAGCAGAGCGCTTATACAGCGCAGCGCTCTCCGCGCAGCCACA 219
Db 392 ACGCAGCTCATGGAGAACAGAGTAGCAATTCACCCAGCAACAGAAATGGATCTCTCAG 451

```

```

QY 220 CAGACAGATGATGCGCGCCCGACCGG---CCAGCCCCCAGACACACCTTCTCAAAC 276
Db 452 CAGACAGAGGTGGAGCACAGACAGCGGCAACAGTACAGACACAAAGTAGTGAAT 511
QY 277 ACAGAAACAGTCCAGCCCAAGCGGTGATGTGTCAACATCCCTTCCGGTTCGG 336
Db 512 TCAGAGAGTAATCTACGCCCAAGGACTACATGTCTCTAATATTCCTTCCGCTCCGA 571
QY 337 GATCCAGACCTCCGACAAATGTTTGGCCAAATTTGGTAAATATTAGATGTTGAAATATT 396
Db 572 GACCCTGACCTCCGCGAGATGTTGGCAGATTGGGCAAAATCTAGATGTGGAATAATC 631
QY 397 TTTAATGACGGGCTCGAAGGATTTGGTTTCGTAACCTTCGAAATAAGTCGGATGG 456
Db 632 TTTAATGACGGGCTCCAAAGGATTCGCGGTTTCGTAACCTTCGAAATAAGTCGATGCA 691
QY 457 GACAGGCGCAGGAGAAATTTGACCGGTACCGTACCGTGTAGAGGCGGTAAATTCGAGGTTAAT 516
Db 692 GACAGGCGCAGGAGAAATTTGACCGCACCGTGTAGAGGCGGTAAATTCGAGGTTAAT 751
QY 517 AATGCGACAGCAGCGGTGATGACTAATAAAGGCGGTGAAACCCCTACACCAATGGCTGG 576
Db 752 AATGCAACAGCAGCGGTGATGACTAATAAAGGCGGTGATGATGTCAGGAAGAGGATC 811
QY 577 AAATTAATCCAGTTGTGGCGCGGTCTACAGCCCGGCTTCTATCGAGCAGCGGTGCT- 635
Db 812 AAGTTAAGCCAGTAGTGTGAGCTGTATGTCCTCGATTTATATGACGATCCAGCTTT 871
QY 636 -----GTTGTCCAGGCGCAACAGAGGAGTCTTCCATGTACAGTGGCCCGGTTCA 687
Db 872 CAAGCTGATGTGCTTCCATAGGCAATGAGGCGGTGCTTATGTCAGGAAGAGGATC 931
QY 688 CTGTGTATATCTTCTGCA---ATGCTGTGCTTCCATATC---CGGCGGCACTGCTGCA 741
Db 932 AACACTTACATCTCTAATCATTTCTGCTGCTTCCCTTACCCAACTGGAGCCACACGCA 991
QY 742 GCTGCATACCGAGGCTACCTTCGAGGCGGTGTCGACCGCTGTACACACCTTCAGA 801
Db 992 GCTGCTTTCGAGAGGAGCCCATCTGAGGGGAGAGGCGGACAGTGTATGTCAGTCCGA 1051
QY 802 GCTCGGCGGCGCCACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCCAGTG 861
Db 1052 ---GCGGTACCTCCAAACAGCATCCCGCTATCCAGGTGTGGTTTACCAGGACGGATT 1108
QY 862 TATGGCAATAATTGCTACAGGCTGTACGCTGATACGCTACGCTACGCTACGCTACGCT 921
Db 1109 TACGGTGTGTA---CCTCTATGTTGGATATGACGCTTACAGATATGACAGCTGTACT 1165
QY 922 GCCA-----CTGCTGCTGCTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 957
Db 1166 GCAACCGCAGCAGCTGCTGACGCGCTGACGCGCTTACAGCGAGGTTACGGCAGG 1225
QY 958 GTTATGCTGCGACCCCTACACACACACTTGTCTCAGCCCCCCTACGCGGCTGGT 1017
Db 1226 GTGTACAGCTGACCCCTACCATGCC---CTCGCCCTGCCCGCAGCTATGGAGTTGGC 1282
QY 1018 GCCATG 1023
Db 1283 GCTGTG 1288

RESULT 15
AC120661/c 115860 bp DNA linear HTG 23-JUL-2002
LOCUS Rattus norvegicus clone CH230-34c2, *** SEQUENCING IN PROGRESS ***
DEFINITION 21 unordered pieces.
ACCESSION AC120661
VERSION AC120661.2 GI:21902877
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```
/product="hexaribonucleotide binding protein 1 isoform
alpha"
/protein_id="AAL83405.1"
/db_xref="GI:19032366"
/translations="MLASQGVLLHPHYGVPMIVPAAPYLPGLIQNQEAAPDPTMAQP
YASQAPAPONGIPAEYTPAPHPHAPPEYTGQTTPEHTLNLXPAPQTHSEQSPADTSA
QTVSGTATOTDDAAPTGGOPOTOPSENKQKRLHVSNI PFRFRDPLRQMFQGF
GKILVVEIIFNERSGKRGFPVTEENSADADRAKELHGTVEGRKLEVNATRVMTN
KKTVPYTNKUNLPVVGAVYSPFAGTVLLCOANQEGSSMYSAPESSIVYTSAMPGF
PYPAATAAPYGAHLRGRRTVINFRAPAAPPIPAYGGVYQBPVYGNKLQGGY
AAYRYAQPTPATAYSDSYGRVYADPVYHALAPATYTVGVGAMASITRGGINRFAPY
"
BASE COUNT      380 a      448 c      374 g      345 t
ORIGIN
Query Match      59.6%; Score 798.4; DB 9; Length 1547;
Best Local Similarity 89.9%; Pred. No. 7e-206;
Matches 935; Conservative 0; Mismatches 96; Indels 9; Gaps 7;
Qy 1 GCGGCGCCCTGACACAAATGCTCAG-CTTATGCTCAGCGCAGTTCGCTCCACCCCGAG 59
Db |||
Qy 356 GCGGCTGCGCCCTGACACAAATGCTCAGCTTACGCTTCGGCCCGAGTTGCTCCCGCGAG 415
Db |||
Qy 60 AATGCAATCTT-GCAGAATACACGCGCCCTCATCCCATCCCGCGCCAGAG-ACACGCGC 117
Db |||
Qy 416 AACGGTATCCCGCGGATACACGCGCCCTCATCCCATCCCGCGCCAGAGTACACAGGC 475
Db |||
Qy 118 CAGCCCATGTCGCCG-CCACACATTAACCTTG-ATCCTCTTACACAGCCGACCTCCGAG 175
Db |||
Qy 476 CAGACCGCGTTCCCGAGCAGACATTAACCTTGATACCTTCGCGCCGCGCAGCTCCGAG 535
Db |||
Qy 176 CAGACG---GCTTATACAGCGCACAG-CGGTCTCCGCGCAGCCGCGCACACAGACATGAT 231
Db |||
Qy 536 CAGAGCCCGCGGACACAGCGCTCAGACCGTCTCTGGCAGCCGCGCACACAGACATGAC 595
Db |||
Qy 232 GCGGCGCCGACGCGCGCGCCAGCACACAACTTCTGAAAAACACAGAAACAAAGTCC 291
Db |||
Qy 596 GCAGCACCGCGAGTGTGTCACACACCACTTCTGAAAAACACAGAAACAAAGTCT 655
Db |||
Qy 292 CAGCCCAAGCGGTGATGTGTCACACACCACTTCTGAAAAACACAGAAACAAAGTCC 351
Db |||
Qy 656 CAGCCCAAGCGGTGATGTGTCACACACCACTTCTGAAAAACACAGAAACAAAGTCC 715
Db |||
Qy 352 CAAATGTTGGCAATTTGCTAAATATTAGATGTTGAAATATTATTTTAAATGAGCGGGC 411
Db |||
Qy 716 CAAATGTTGGTCAATTTGCTAAATATTAGATGTTGAAATATTATTTTAAATGAGCGGC 775
Db |||
Qy 412 TCAGAGGATTTGGTTTCGTAATTTTCGAAATATGTCGGATCGCGACAGGCGGAGGAG 471
Db |||
Qy 776 TCAGAGGATTTGGTTTCGTAATTTTCGAAATATGTCGGATCGCGACAGGCGGAGGAG 835
Db |||
Qy 472 AAATTCACGGTACCGTGTAGAGCGCGTAAATAGAGTTAAATAGAGTTAAATAGAGTACG 531
Db |||
Qy 836 AAATTCACGGTACCGTGTAGAGCGCGTAAATAGAGTTAAATAGAGTTAAATAGAGTACG 895
Db |||
Qy 532 GTGATGACTAATAAAGGCGGTGAACCCCTACACCAATAGGCTGGAATTTAAATCCAGTT 591
Db |||
Qy 896 GTAAATGACAAATAAAGAGCCGTCAACCTTTATACAAATAGGCTGGAATTTAAATCCAGTT 955
Db |||
Qy 592 GTGGGCGCGTCTACAGCCCGCATTTCTATGAGGACAGCGTGTGTTGTCGAGGCGAAC 651
Db |||
Qy 956 GTGGGTCAGTCTACAGTCCGAAATTTCTATGAGGACAGCGTCTGTTGTCGAGGCGAAC 1015
Db |||
Qy 652 CAGGAGGATCTTCATGATACAGTGCGCCCGCATTTCTATGATATCTTCTGCAATGCCT 711
Db |||
Qy 1016 CAGGAGGATCTTCATGATACAGTGCGCCCGCATTTCTATGATATCTTCTGCAATGCCA 1075
Db |||
Qy 712 GGCTTTTCATATCGGCGCGCATGCTGTCAGCTGCATACAGAGGGGCTCACTTCGAGGC 771
Db |||
Qy 1076 GGCTTCCGATCCAGAGAGCCACCGCGCGCGCTTACAGAGGGGCGACCTTCGAGGC 1135
Db |||
Qy 772 CGTGGTCCGACCGGTGTAACACCTTCAGAGCTCGGCGCGCCACCCCAATCCCGGCC 831
Db |||
```

```
Db 1136 CCGGTGCGCACCGTGTACAAACACCTTTCAGGCGCGCGCGCGCGCGCGCGCGCGCC 1195
Qy 832 TATGCGGAGTAGTGTATCAAGAGCCAGTGTATGCAATAAATTTCTACAGGCTGTTAC 891
Db |||
Db 1196 TACGCGGAGTAGTGTATCAAGAGCCAGTGTATGCAATAAATTTCTGCGAGGCTTAC 1255
Qy 892 GCTGATACCGTACGCGCCAGCCACCCCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 951
Db |||
Db 1256 GCTGATACCGTACGCGCCAGCCAGTGTATGCAATAAATTTCTGCGAGGCTTAC 1315
Qy 952 GGACGAGTTTATGCTGCGCCAGCCCTTACACACACACTTGTCTCCAGCGCCCTACCGC 1011
Db |||
Db 1316 GGACGAGTTTATGCTGCGCCAGCCCTTACACACACACTTGTCTCCAGCGCCCTACCGC 1375
Qy 1012 GTTGGTGCATGAATGCTTT 1031
Db |||
Db 1376 GTTGGTGCATGCTAGTAT 1395
Qy |||
Db |||
RESULT 12
HSM803066 1623 bp mRNA linear PRI 20-MAR-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp547L059 (from clone DKFZp547L059).
DEFINITION AL713700
ACCESSION AL713700.1 GI:19584415
VERSION
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1623)
AUTHORS Bloeker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B. and
Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp547L059) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..1623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="16p13.3"
/clone="DKFZp547L059"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbri). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="fetal"
1..1623
/gene="DKFZp547L059"
836..>1623
/gene="DKFZp547L059"
/note="FLJ10165 fis"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD28499.1"
/db_xref="GI:19584416"
/translation="MNCERQLRGNQEAAPDPTMAQFYASQAPQNGIPAEYTPAP
HPHAPPEYTGQTTPEHTLNLXPAPQTHSEQSPADTSAQTVSGTATOTDDAAPTGGOP
QTVSGTATOTDDAAPTGGOPOTOPSENKQKRLHVSNI PFRFRDPLRQMFQGFQGF
GKILVVEIIFNERSGKRGFPVTEENSADADRAKELHGTVEGRKLEVNATRVMTNKTVPY
TNKUNLPVVGAVYSPFAGTVLLCOANQEGSSMYSAPESSIVYTSAMPGFPYPAATA"
BASE COUNT      356 a      523 c      487 g      257 t
ORIGIN
```

```
/gene="HRNBP1"  
/note="similar to Caenorhabditis elegans fox-1"  
/codon start=1
```



```

gene      1..2279
          /gene="A2Bp"
CDS       1047..2180
          /gene="A2Bp"
          /codon_start=1
          /product="ataxin 2-binding protein"
          /protein_id="AAF78291.1"
          /db_xref="GI:8671586"
          /translation="MAQPYASAGAPPONGIPAEYTAHPHPAPEYTGQTTVPEHTLN
LYPQAHTSESPADTSMAOTVSGTATDDBAATDGOPTQPSQENTENKSPQKELHVS
NIPRFRPDLRQMGFGKTLIDVEIIFNERSKGFGFTFENSADARAKLHGTV
VEGKIEVNNATARNNNKTNVNYTWGKLNVPVGVAVISPEFYAGTVLLCQANQGS
SMYSPASLVTSAMPQFPYPAATAAAYGAHLRGRGRTVYNTFRAAAPPPPIPAYG
GVTVQDFYGIADYGGYAAVRYAQTPTATAAASDSYGRVYAADPVYHALAPARTYGV
GAMNAFPLTDKAKTRSHADDVGLVLSLQASIVRGYNNRPAY"
          /misc_feature
          /gene="A2Bp"
          /note="Region: RNA binding motif RNP-2"
          /misc_feature
          /gene="A2Bp"
          /note="Region: RNA binding motif RNP-1"

BASE COUNT 497 a 712 c 636 g 434 t
ORIGIN

Query Match      67.5%; Score 904.2; DB 9; Length 2279;
Best Local Similarity 86.7%; Pred. No. 1.2e-234;
Matches 1117; Conservative 0; Mismatches 123; Indels 49; Gaps 9;

Oy 1 GCGGCGGCCCTGACACAATGCCTCAG-CTTATGCCTCAGCGCAGTTCGCTCCACCCAG 59
Db 1029 GCCGTGCCCCTGACACAATGCCTCAGCCTTACGCTTCGGCCCCAGTTGCTCCCCCGAG 1088

Oy 60 AATGCACTCCT- GCAGAAATACACGCGCCCTCATCCCCATCCCGGCCGACAG-ACACCGGC 117
Db 1089 AACGGTATCCCGCGGAATACGCGCCCTCATCCCAACCCCGGCCGACAGTACACAGGC 1148

Oy 118 CAGGCCACTGTCGCG- CCACACATTAACATTG- ATCTCTCTACACAGACGCACTCGGAG 175
Db 1149 CAGACCACGGTTCGCGAGCACACATTAACCTGTACCTTCGCGCCGACGCACTCCGAG 1208

Oy 176 CAGAGC---GCTTATACAAGCGCACAG- CCGTCTCCGGCACCCGCCACACAGACATGAT 231
Db 1209 CAGAGCCCGCGGACACGAGCGCTCAGACCGCTCTCTGGCACCCGCCACACAGACATGAC 1268

Oy 232 GCCGCCCGCAGCGGCAGCGCCGACACACAACTTCTGAAAAACACAGAAAAACAAGTCC 291
Db 1269 GCAGCAGCGAGTGCAGCGCCGACGACACACACTTCTGAAAAACACGGAACAAGTCT 1328

Oy 292 CAGCCCAAGCGGCTGCATGTGTCCAAATCCCTTCGGGTTCCGGATCCAGACCTCCGA 351
Db 1329 CAGCCCAAGCGGCTGCATGTGTCCAAATATCCCTTCAGGTTCCGGATCCGGACCTCAGA 1388

Oy 352 CAAATGTTTGCCCAATTTGGTAAATATTAGACTTGGAAATTAATTTTATGAGCGGGC 411
Db 1389 CAAATGTTTGGTCAATTTGGTAAATCTTAGATGTTGAAATTAATTTTATGAGCGAGGC 1448

Oy 412 TCGAAGGATTTGTTTTCGTAACCTTTTCGAAAAATAGTCGGATCGGCACAGGGCGAGGGAG 471
Db 1449 TCAAAGGATTTGGTTTTCGTAACCTTTTCGAAAAATAGTGCCGATCGGCACAGGGCGAGGGAG 1508

Oy 472 AATTTGACGGTACCGTGTGTAGAGGCGGTAAATCGAGGTTTAATTAATCGCACAGCACGC 531
Db 1509 AATTTACCGGCAACCGTGTGTAGAGGCGGTAAATCGAGGTTAAATTAATGCCACAGCACGT 1568

Oy 532 GTGATGACTAATAAAAAGCGCGTGAACCCCTACACCAATGGCTGGAAATTAATTCAGTT 591
Db 1569 GTAATGACAAATAAAAAGACCGTCAACCCCTTATACAAATGGCTGGAAATTAATTCAGTT 1628

Oy 592 GTGGGCGCGGTCTACAGCCCGGACTTCTATGACGACCGGTGCTGTGTGTCAGGCGCAAC 651
Db 1629 GTGGGTGCACTACAGTCCCGAATCTATGACGACCGGTCTGTGTGTCAGGCGCAAC 1688

Oy 652 CAGAGGGATCTTCATGTACAGTGCGCCCGAGTTCATTTGTATATATCTTCTGCAATGCT 711

```

QY	652	CAGGAGGGATCTTCCATGTACAGTGCGGCCCACTTCCTGTATATACTTCTCGAATGCCT	711
Db	923	CAGGAGGGATCTTCCATGTACAGTGCGGCCCACTTCCTGTATATACTTCTCGAATGCCT	982
QY	712	GGCTTTCCATATCCCGCGCGCATCTGTCAGCTGCATACCGAGGGGCTCACCTTCGAGGC	771
Db	983	GGCTTTCCCGTATCCAGCAGCCACCGCGCGCGCTACTCGAGGGGCGCACCTGCGAGGC	1042
QY	772	CGTGTGCGCACCGGTGTACAACACCCTTCAGAGCTGCGGCGCCCCACCCCACATCCCGGCC	831
Db	1043	CGCGTTCGCAACCGGTGTACAACACCCTTCAGGCGCGCGCGCCCCCGCGATCCCGGCC	1102
QY	832	TATGGCGGAGTAGTGTAACAGAGCCAGGTGTATGGCAATAAAATTGCTACAGGCTGGTTAC	891
Db	1103	TACGGCGGCTGTTGTTTACCAAGGATGATTTATGG---TGCAGACATTTATGGTGGTTAT	1159
QY	892	GCTGCATACCGCTACGCGCCAGCCACCCCTGCCACTGCTGCTGCCTACAGTGACAGTTAC	951
Db	1160	GCTGCATACCGCTACGCGCCAGCCCTACCCCTGCCACTGCGCTGCCTACAGTGACAGTTAC	1219
QY	952	GGACGAGTTTATGTCGCGGACCCCTTACACACACACTTGTCTCCAGGCCCCACCTACGGC	1011
Db	1220	GGACGAGTTTATGTCGCGGACCCCTTACACACAGCACTTGTCTCCAGGCCCCACCTACGGC	1279
QY	1012	GTTGGTGCCTATGAATGCTTTTGGCGCCCTTCACCGATGCCAAGACTAGGAGCCATGCTGAT	1071
Db	1280	GTTGGTGCCTATGAATGCTTTTGGCGCCCTTCACCGATGCCAAGACTAGGAGCCATGCTGAT	1339
QY	1072	GATGTGGTCTCGTGTCTTTCTTTCATTCAGGCTAGTATATACCAAGGGGATACCAACCGT	1131
Db	1340	GATGTGGTCTCGTGTCTTTCTTTCATTCAGGCTAGTATATACCAAGGGGATACCAACCGT	1399
QY	1132	TTTGTGCTATTAATATGAATAAACCACTTAACAAAACAAAGCAAAAAACAAAAACA	1191
Db	1400	TTTGTGCTATTAATATGAATAAACCACTTAACAAAACCAATAA-----	1430
QY	1192	AAAAAACCAACCTTCCAATGTGGGAGAGAGAAGCTTTCGAGGCCGAGTGTGGCAC	1251
Db	1431	-----AACCTTCCAATGTGGGAGAGAAGAGCTTTCGAGGCCCTGAGTATTGCAAT	1482
QY	1252	ACATGCATGAGACATCATCTTTAGCAACTC	1281
Db	1483	ACATGCATGAGACATCATCTTTAGCAACTC	1512
RESULT 8			
AF107203		AF107203	2279 bp mRNA linear PRI 23-JUN-2000
LOCUS		Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.	
DEFINITION		AF107203	
ACCESSION		AF107203	
VERSION		AF107203..	GI:8671585
KEYWORDS			
SOURCE		Homo sapiens.	
ORGANISM		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 2279)	
AUTHORS		Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.	
TITLE		A novel protein, A2bp, with RNA binding motif binds to C-terminal ataxin-2	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 2279)	
AUTHORS		Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.	
TITLE		Direct Submisson	
JOURNAL		Submitted (16-NOV-1998) Division of Neurology and Neurogenetics Laboratory, Burns and Allen Research Institute,, Cedars-Sinai Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA	
FEATURES		Location/Qualifiers	
source		1..2279	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/note="similar to Homo sapiens cDNA clone HSP38A20"	


```
LYPTOTHSQSADTSAGTUSGTATDADPTDQDQOTOPSENTESSKOPKRLHVSNI
IPFRFRDPLRQMFQKGLDVEIIFNERSGKGFVTFENSADADRARLKHGT
VBERKLEVNATRVMTNFKVNPVNGKLVNPGVAVSPDFYAGTVLLCOANOBG
SSMTSGSSSLVYTAMPGFPYPAAATAAAVGAHRLGRGRTVNTFRAPAAPBPPIAY
GGVYQDGFYADLYGGVAAVYAAQPTPATAAAVSDSYGRVYAADPVHHTLAPPTYG
VGMNAPAPLTDKTRSHADVDGLVLSLQNSIYRGGINRFAPY"
misc_feature
358..375
/gene="A2bp"
/note="Region: RNA binding motif RNP-2"
475..495
/gene="A2bp"
/note="Region: RNA binding motif RNP-1"
BASE COUNT 362 a 396 c 324 g 281 t
ORIGIN
Query Match 80.9%; Score 1084.4; DB 10; Length 1363;
Best Local Similarity 93.6%; Pred. No. 1e-283;
Matches 1233; Conservative 0; Mismatches 66; Indels 19; Gaps 9;
QY 1 GCGCGCGCCCTGACACAAATGGCTCAG-CTTATGCTCAGCGCAGTTCGCTCCACCCCGAG 59
DB 49 GCGCGCGCCCTGACACAAATGGCTCAGCTTATGCTCAGCGCAGTTCGCAACCCCGAG 108
QY 60 AATGCGCAT-CTGCGAATACACGCGCCCTCATCCCATCCGCGCCGAGAG-ACACCGCG 117
DB 109 AATGCGCATCTGAGAAATACACGCGCCCTCATCTCATCCGCGCCGAGAGTACACCGCG 168
QY 118 CAGCCCACTGTCCTCG-CCACACATTAACTTG-ATCCTCTCACAGAGCGCACTCGGAG 175
DB 169 CAGACCACTGTCCTCGACACACATTAAACCTGTATCTCTCTACAGAGCGCACTCGGAG 228
QY 176 CAGAGCCCTTATACAGGGACAG-CGCTCTCGGACCGCCACACAGACAGATGATGCC 234
DB 229 CAGAGTGTGACACAGTGGCGACAGCGTCTCGGACCGCCACACAGACAGATGATGCA 288
QY 235 GCCCGGACCGAGCGCCGACACACAACTTCTGAAACACAGAAACAGTCCCGAG 294
DB 289 GCCCGGACCGAGTGGCGAGCCGACACAACTTCTGAAACACAGAAACAGTCCCGAG 348
QY 295 CCAAGCGGTGATGTGTCAACATCCC-----CTTCGCGTTCGCGGATCCGAGACCTC 348
DB 349 CCAAGCGGTGATGTGTCAACATCCCCTTCGCGTTCGCGGATCCGAGACCTC 408
QY 349 CGACAAATGTTTGGCCAAATTTGTAATATATAGATGTTGAAATTTATTAATAGCGG 408
DB 409 CGACAAATGTTTGGTCAATTTGTAATATATAGATGTTGAAATTTATTAATAGCGCA 468
QY 409 GGCTCGAAGGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGATCGGACGGCGAGG 468
DB 469 GGCTCGAAGGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGATCGGACGGCGAGG 528
QY 469 GAGAAATTGACGGTACCGGTGAGAGGCGGTAATAATCGAGTTAATAATGCGACAGCA 528
DB 529 GAGAAATTGACGGTACCGGTGAGAGGCGGTAATAATCGAGTTAATAATGCGACAGCA 588
QY 529 CGCGTGATGACTAATAAAGGCGGTGAACCCCTCACCAATGGCTGGAATAATAATCCA 588
DB 589 CGCGTGATGACTAATAAAGGCGGTGAACCCCTCACCAATGGCTGGAATAATAATCCA 648
QY 589 GTTGTGGCGCGGTCTACAGCCCCGACTTCTATGAGGACCGTGTGTTGTGCCAGGCC 648
DB 649 GTTGTGGCGCGGTCTACAGCCCCGACTTCTATGAGGACCGTGTGTTGTGCCAGGCC 708
QY 649 AACCAGAGGGATCTTCATGTACAGTGGCCCGAGTTCATTTGATATATCTTCGCAATG 708
DB 709 AACCAGAGGGATCTTCATGTACAGTGGCCCGAGTTCATTTGATATATCTTCGCAATG 768
QY 709 CTGTGCTTTCCATATCCGCGCCGCTGTCAGCTGCATACAGAGGGGCTCACCTTGA 768
DB 769 CTGTGCTTTCCATATCCGCGCCGCTGTCAGCTGCATACAGAGGGGCTCACCTTGA 828
QY 769 GCGCGGTGTCGACCGGTGTACAACTTTCAGAGTTCGCGCGCCCGCCACCCCAATCCG 828
```

```
Db 829 GCGCGTGGTCGACCGGTGTAAACACACTTCAGGGCTGCGAGCGCCCGCCCAATCCCG 888
QY 829 GCTATGGCGGAGTACTGTATCAAGAGCCAGTGTATGGCAATAAATTTGTACAGGGTGGT 888
Db 889 GCTATGGCGGAGTGTGTATACCGAGATGATTTATGG---TGACAGACATTTATGGTGGT 945
QY 889 TAGCGTGCATACCGCTACGCGCCAGCCGACCCCTGCGCACTGCTGCTCCCTACAGTGACAGT 948
Db 946 TATGCTGGCTACCGCTATGCCAGCGCCACCCCTGCGCACTGCGCTGCCCTACAGTGACAGT 1005
QY 949 TAGCGACGAGTTTATGCTGCCGACCCCTACCAACACACTTTGCTCCAGCCCCCAGCTTAC 1008
Db 1006 TAGCGACGAGTTTATGCTGCCGACCCCTTACCACACACACTTGTCTCCAGCCCCCAGCTTAC 1065
QY 1009 GCGTGTGGTGGCCATGATGCTTTTGGCGCCCTTGACCGATGCCAAGCTAGGAGCCATGCT 1068
Db 1066 GCGTGTGGTGGCCATGATGCTTTTGGCGCCCTTGACCGATGCCAAGCTAGGAGCCATGCT 1125
QY 1069 GATGATGTGGTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1128
Db 1126 GATGATGTGGTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1185
QY 1129 CGTGTGTGCTCCATATTAATGATAAAACCATTAACAAACAAACAAACAAACAAACAAACAAAC 1185
Db 1186 CGTGTGTGCTCCATATTAATGATAAAACCATTAACAAACAAACAAACAAACAAACAAACAAAC 1245
QY 1186 -AAAAACAAACAAACCAACCTTCCAAATGTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
Db 1246 GAAACAAACAAACCAACCTTCCAAATGTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
QY 1245 TTGCGACACATGACGTAGGACATCACTTTAGCAACTCAAGAAACAAACGAAAAACAAACAAAC 1302
Db 1306 TTGCGATCATGACGTAGGACATCATGTTAGCAACTCAAGAAACAAACGAAAAACAAACAAAC 1363
```

```
RESULT 5
AF229057
LOCUS Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
DEFINITION AF229057 2000 bp mRNA linear PRI 01-FEB-2002
ACCESSION AF229057 (HRNBPI) mRNA, complete cds.
VERSION AF229057.1 GI:18461366
KEYWORDS SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (Bases 1 to 2000)
AUTHORS Chen, W. and Winkelman, J.C.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology, for
University of Cincinnati College of Medicine, The Vontz Center
Molecular Studies, 125 Eden Avenue, Cincinnati, OH 45267-0508, USA
FEATURES
source
1. 2000
/organism="Homo sapiens"
/db xref="taxon.9606"
/chromosome="16"
/map="16p13.3"
1. 2000
/gene="HRNBPI"
256..1512
/notes="RNA-binding protein"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
gamma"
protein_id="AAL71904.1"
/db xref="GI:18461367"
/translaton="MLASQVLLHPVGVPMIVPAAAPLPLIGNQEQEAAAAADPTMAQP
YASQAPAPONGIPAEYTAHPHAPPEYTGQTTVEHTLNLIPPAQTHSEQSPADTSA
QTVSGTATQTDNAPTDGGQTPQSPENKNSOPKRLHVSNIIPFRDPLRQMFQF
GKILDVEIIFNERSGKGFVTFENSADADRARLKHGTIVVNNATRVMTN
```



```
/dev_stage="adult"
637..1827
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA95079.1"
/db_xref="GI:7670456"
/translation="MNCEREQLRGNQAAAAADPTMAQPVASQAQFAPPONGIPAEYVAP
HHPAPEYTGQTPVGHVNIIPPTOTHSQASADTSQVSTATCTDDAQTDSQPO
TOPSENTESKQPKRUHVSNIIPRRPDLROMFGQFGKILLVEIIPNERSKRGFV
TPENSADADAREKLGHTVEGRKLIVANNATARVINKTVNPYINGWKLNPVGVAVY
SPDYAGTVLLCOANGEGSSMYGSESSLYTTSAMPGFPYPAATAAAAAGHALLRGRG
TYVNTFRRAAAPPPIPAYGVVYQDFYAGDIYGGYAYRYAQPPTATAAAYSISYGR
VYAADYHHHTLAPATPYGVGNANAPLTDKTRGHADDVGLVLSLQASIVRGYNR
FAPY"
BASE COUNT      518 a   589 c   516 g   379 t
ORIGIN

Query Match      82.8%; Score 1110; DB 10; Length 2002;
Best Local Similarity 94.1%; Pred. No. 1.2e-290;
Matches 1242; Conservative 0; Mismatches 65; Indels 13; Gaps 8;

QY 1 GCGGCGCCCTGACACATGGCTCAG-CCTATGCTCTCAGCGAGTTCGCTCCACCCCGAG 59
DB 679 GCGGCGCCCTGACACATGGCTCAGCTTATGCTCTCAGCGAGTTCGACACCCCGAG 738
QY 60 AATGGCAT-CTTCGAGAAATACAGCGCCCTCATCCCATCCCGCGCCAGAG-ACACCGCG 117
DB 739 AATGGCATCCCTGACAAATACAGCGCCCTCATCCCTCATCCCGCGCCAGAGTACACCGCG 798
QY 118 CAGCCCACTGTCCC-CCGCACACATTAACCTTG-ATCCCTCCTACACAGACGACTCGGAG 175
DB 799 CAGACCACTGTCCCTGCCACACATTAACCTTGATCCCTTACACAGACGACTCGGAG 858
QY 176 CAGAGCGCTTATACAGCGCCAG-CGCTCTCCGCGCACCGCCACACACAGACATGATGCC 234
DB 859 CAGAGTGTCTGACACAGTGGCAGACCGCTCCGCGCACCGCCACACAGACATGATGCA 918
QY 235 GCGCCGACCGCGCCAGCGCCAGACAGAACTTCTGAAAAACAGAGAAACAAAGTCCCGAG 294
DB 919 GCGCCGACCGCGCCAGCGCCAGACAGAACTTCTGAAAAACAGAGAAACAAAGTCCCGAG 978
QY 295 CCCAAGCGCTGCATGTGTCCAACTTCCCGTTCGCGGATCCGAGCTCCCGACAA 354
DB 979 CCCAAGCGCTGCATGTGTCCAACTTCCCGTTCGCGGATCCGAGCTCCCGACAA 1038
QY 355 ATGTTGGCAATTTGGTAAATATATAGATGTGAAATATATTTTAAATCAGCGGGCTCG 414
DB 1039 ATGTTGGTCAATTTGGTAAATATATAGATGTGAAATATATTTTAAATCAGCGGGCTCC 1098
QY 415 AAGGATTTGGTTTCGTAACTTTCGAAATAGTGGGATCGGACAGCGCGGAGAGAA 474
DB 1099 AAGGATTTGGTTTCGTAACTTTCGAAATAGTGGGATCGGACAGCGCGGAGAGAA 1158
QY 475 TTGCAAGGTACCGTGTAGAGCGCGTAAATTCGAGGTAAATTCGACACAGCAGCGGTG 534
DB 1159 TTGCAAGGTACCGTGTAGAGCGCGTAAATTCGAGGTAAATTCGACACAGCAGCGGTG 1218
QY 535 ATGACTTAATAAAGCGCGTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 594
DB 1219 ATGACAAATAGAGACATGTCAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 1278
QY 595 GCGCGGTCTACAGCCCGACTTCTATCAGCGACGGTGTGTTGTGCGAGGCCAACCGAG 654
DB 1279 GCGCGGTCTACAGCCCGACTTCTATCAGCGACGGTGTGTTGTGCGAGGCCAACCGAG 1338
QY 655 GAGGATCTTCATGTACAGTGGCCCGAGTTCACTTGTATATATCTTGTCAATGCTGGC 714
DB 1339 GAGGATCTTCATGTACAGTGGCCCGAGTTCACTTGTATATATCTTGTCAATGCTGGC 1398
QY 715 TTTCCATATCCGGCGGCACCTGCTGACGTCATACCGAGGGCTCACCTTCGAGGCGGT 774
DB 1399 TTTCCATATCCGGCGGCACCTGCTGACGTCATACCGAGGGCTCACCTTCGAGGCGGT 1458
```

```
QY 775 GGTCCACCGTGTACACACACTTCAGAGCTCGGCGCCCCCACCACCCCAATCCCGGCTTAT 834
DB 1459 GGTCCACCGTGTACACACACTTCAGAGCTCGGCGCCCCCACCACCCCAATCCCGGCTTAT 1518
QY 835 GCGGAGTAGTGTATCAAGAGCCAGTGTATGCAATAAATGCTACAGGGTGTAGCT 894
DB 1519 GCGGAGTAGTGTATCAAGAGTGTATGCAATAAATGCTACAGGGTGTAGCT 1575
QY 895 GCATACCGCTACGCCACGCCACCCCTGCCACTGCTGCTACAGTACAGTACAGTACCGA 954
DB 1576 GGTACCGCTATGCCAGCCACCCCTGCCACTGCTGCTACAGTACAGTACAGTACCGA 1635
QY 955 CGAGTTATGCTGCGGACCCCTACACACACACTTCTCAGCCCCCAGCTACGGGCTT 1014
DB 1636 CGAGTTATGCTGCGGACCCCTACACACACACTTCTCAGCCCCCAGCTACGGGCTT 1695
QY 1015 GGTGCCATGAATGCTTTTGGGCCCTTGACCGATGCCAAGACTAGGAGCCATGCTGATGAT 1074
DB 1696 GGTGCCATGAATGCTTTTGGGCCCTTGACCGATGCCAAGACTAGGAGCCATGCTGATGAT 1755
QY 1075 GTGGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1134
DB 1756 GTGGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1815
QY 1135 GCTCCATATTAAATGATAAAACCAATTAAACAAACAGCAAAACAAACAAACAAACAA 1190
DB 1816 GCTCCATATTAAATGATAAAACCAATTAAACAAACAGCAAAACAAACAAACAAACAA 1875
QY 1191 AAAAAACCAACCTTCAATGTGGGAGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGCA 1250
DB 1876 AAAAAACCAACCTTCAATGTGGGAGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGCA 1935
QY 1251 CACATGAGTAGACATCCTTTAGCACTCTAAAGAAACAAACAAACAAACAAACAAACAA 1310
DB 1936 CACATGAGTAGACATCCTTTAGCACTCTAAAGAAACAAACAAACAAACAAACAAACAA 1995
```

```
RESULT 4
AF107204 LOCUS 1363 bp mRNA linear ROD 23-JUN-2000
DEFINITION Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.
JOURNAL A novel protein, A2BP, with RNA binding motif, binds to C-terminal
REFERENCE ataxin-2
AUTHORS Unpublished
TITLE 2 (bases 1 to 1363)
JOURNAL Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.
REFERENCE Direct Submission
AUTHORS Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
TITLE Laboratory, Burns and Allen Research Institute, Cedars-Sinai
JOURNAL Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
FEATURES Angeles, CA 90048, USA
source
1. .1363
/organism="Mus musculus"
/db_xref="taxon:10090"
gene
1. .1363
/feature="A2bp"
CDS
67..1203
/gene="A2bp"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AAF78292.1"
/db_xref="GI:8671588"
/translation="MAQPVASQAQFAPPONGIPAEYVAPHPHAPAEYTGQTTVPDHTLN
```


JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;

Scios Inc. (US)

Location/Qualifiers

FEATURES

source

1. .1340

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

BASE COUNT 366 a 394 c 317 g 263 t

ORIGIN

Query Match 100.0%; Score: 1340; DB 6; Length 1340;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGCGCCCTGACCAATGGCTCAGCTTATGCTCAGCGCAGTTTCGCTCCACCCCGA 60
DB 1 GCGGCGCCCTGACCAATGGCTCAGCTTATGCTCAGCGCAGTTTCGCTCCACCCCGA 60

QY 61 ATGGCATCTGACAGATACACGGCCCTCATCCCATCCCGCCAGACACCGCCGAG 120
DB 61 ATGGCATCTGACAGATACACGGCCCTCATCCCATCCCGCCAGACACCGCCGAG 120

QY 121 CCACATGTCCTCCGCCACACATTAACCTTGATCTCTACACAGACGACCTCGAGCAGAG 180
DB 121 CCACATGTCCTCCGCCACACATTAACCTTGATCTCTACACAGACGACCTCGAGCAGAG 180

QY 181 CGCTTATACAGCGCAGCGCTCCGGACCGCCACACACAGATGATGCCGCCCG 240
DB 181 CGCTTATACAGCGCAGCGCTCCGGACCGCCACACACAGATGATGCCGCCCG 240

QY 241 ACCGACGGCCAGCCCGACACAACTCTCTGAAACACAGAAACAAAGTCCAGCCCAAG 300
DB 241 ACCGACGGCCAGCCCGACACAACTCTCTGAAACACAGAAACAAAGTCCAGCCCAAG 300

QY 301 CGGCTGCATGTGTCACATCCCTTCGGTTCCGGATCCAGACCTCCGACAAATGTTT 360
DB 301 CGGCTGCATGTGTCACATCCCTTCGGTTCCGGATCCAGACCTCCGACAAATGTTT 360

QY 361 GCGCAATTTGGTAAATATAGATGTTGAATATTTTAAATGAGCGGGCTCGAAGGGA 420
DB 361 GCGCAATTTGGTAAATATAGATGTTGAATATTTTAAATGAGCGGGCTCGAAGGGA 420

QY 421 TTTGGTTTCGTAATTTTGGAAATAGTCGGATCGGACAGCGCGAGGAGAAATTCAC 480
DB 421 TTTGGTTTCGTAATTTTGGAAATAGTCGGATCGGACAGCGCGAGGAGAAATTCAC 480

QY 481 GGTACCGTGTAGAGGCGGTAATAATCGAGTTAATAATGCGACAGCAGCGTGATGACT 540
DB 481 GGTACCGTGTAGAGGCGGTAATAATCGAGTTAATAATGCGACAGCAGCGTGATGACT 540

QY 541 AATAAAGGCGGTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTGGGCGG 600
DB 541 AATAAAGGCGGTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTGGGCGG 600

QY 601 GTCTACAGCCCGACTTCTATGAGGCGAGTGTGTTGTGCGAGGCCAACAGAGGGA 660
DB 601 GTCTACAGCCCGACTTCTATGAGGCGAGTGTGTTGTGCGAGGCCAACAGAGGGA 660

QY 661 TCTTCCATGTACAGTGGCCCGGCTACCTGTATATCTCTGCAATGCTGCTTCCCA 720
DB 661 TCTTCCATGTACAGTGGCCCGGCTACCTGTATATCTCTGCAATGCTGCTTCCCA 720

QY 721 TATCCGGCGGCGACTGCTGACGCTCATACAGGGGCTCACCTTCGAGGCGGTTGTCG 780
DB 721 TATCCGGCGGCGACTGCTGACGCTCATACAGGGGCTCACCTTCGAGGCGGTTGTCG 780

QY 781 ACCGTGTACACACTTTCAGAGTGGCGGCCCGCCACCCCCCAATCCCGGCTATGGCGGA 840
DB 781 ACCGTGTACACACTTTCAGAGTGGCGGCCCGCCACCCCCCAATCCCGGCTATGGCGGA 840

QY 841 GTAGTGTATCAAGACCGAGTGTATGGCAATAAATGCTACAGGTTGTTAGCTGCTATAC 900
DB 841 GTAGTGTATCAAGACCGAGTGTATGGCAATAAATGCTACAGGTTGTTAGCTGCTATAC 900

QY 901 CGCTAGCGCCGACCCACCCCTGCCACTGCTGCTGCTACAGTGAAGTTCAGGACGAGTT 960
DB 901 CGCTAGCGCCGACCCACCCCTGCCACTGCTGCTGCTGCTACAGTGAAGTTCAGGACGAGTT 960

QY 961 TATGCTGCGGACCCCTACACCAACACACTTGTCTCAGGCCCCCCTACGCGGTTGGTGCC 1020
DB 961 TATGCTGCGGACCCCTACACCAACACACTTGTCTCAGGCCCCCCTACGCGGTTGGTGCC 1020

QY 1021 ATGAATGCTTTTGGCCCTTACCGATGCGCAAGCTAGGAGCCATGCTGATGATGGGT 1080
DB 1021 ATGAATGCTTTTGGCCCTTACCGATGCGCAAGCTAGGAGCCATGCTGATGATGGGT 1080

QY 1081 CTGCTTCTTCTTCTTCAATGCGGCTAGTATATACCAAGGGGATACAAACGTTTGGTCCA 1140
DB 1081 CTGCTTCTTCTTCTTCAATGCGGCTAGTATATACCAAGGGGATACAAACGTTTGGTCCA 1140

QY 1141 TATTAAATGATAAAACCATTAACAAACAAAGCAAAACAAACAAACAAACAAACAAACCA 1200
DB 1141 TATTAAATGATAAAACCATTAACAAACAAAGCAAAACAAACAAACAAACAAACAAACCA 1200

QY 1201 ACCTTCAATGTCGGGAGAGGAGGAGCTTCCGAGGCCGAGTGTTCGGACACATGAGT 1260
DB 1201 ACCTTCAATGTCGGGAGAGGAGGAGCTTCCGAGGCCGAGTGTTCGGACACATGAGT 1260

QY 1261 AGGACATCCTTTAGCAACTCAAGAAACAAACAAACAAACAAACAAACAAACAAACCA 1320
DB 1261 AGGACATCCTTTAGCAACTCAAGAAACAAACAAACAAACAAACAAACAAACAAACCA 1320

QY 1321 GCGCGAAGGGTTCGCTAGA 1340
DB 1321 GCGCGAAGGGTTCGCTAGA 1340

RESULT 2

AF191501

LOCUS

DEFINITION

Mus musculus hexaribonucleotide binding protein 1 (Hrnbp1) mRNA, complete cds.

ACCESSION

AF191501

VERSION

AF191501.1

KEYWORDS

Mus musculus.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 1586)

AUTHORS

Chen, W. and Winkelman, J.C.

TITLE

Direct Submission

JOURNAL

Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, 231 Bethesda Ave., Cincinnati, OH 45267-0508, USA

FEATURES

Location/Qualifiers

1. .1586

/organism="Mus musculus"

/db_xref="taxon:10090"

/tissue_type="skeletal muscle"

gene

1. .1586

/gene="Hrnbp1"

CDS

166. .1419

/gene="Hrnbp1"

/codon_start=1

/product="hexaribonucleotide binding protein 1"

/protein_id="AAL83425.1"

/db_xref="GI:19032414"

/translation="MLASQVLLHSYGVPMIVPAAPYFPGMLQGNQEAAPDPTMAQPVASQAPFQNGIPAEYTAHPHPAPYTGTTGTTVPDHTLNLVPTQTHSEASDTSQTVSGTATDDAAGTGGQPTQPSNTESKQPKRLHVSNIFFRDLRDLQMFQFGKILDEIIFNERSGKGFVTFENSADAREKLHGTVEGRKIEVNATFATVMTNKITVNPYTNKLNPNVGAIVSDFYAGTLLCOANQSGSSMYSGPSSLVYTSAMPFPYPAASAAAYRGKALHGRGRTGYNTLRAAAPPIPAVGVVYQEPVYGNKLQGVAVRVAQPTPATAAAYSDSYGRVYAADPVHHTLAPATYGVGMNAFALPTDAKTRSHADVGVLSSLSQASIVRGYNRPAY"

BASE COUNT 407 a 458 c 374 g 347 t

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:46:57 ; Search time 3583.42 Seconds
(without alignments)
10882.841 Million cell updates/sec

Title: US-09-809-545A-1
Perfect score: 1340
Sequence: 1 gcggcgccctgacacaat.....ggcgaagggttcgctaga 1340

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1340	100.0	1340	6	AX268800	AX268800 Sequence
2	1145.8	85.5	1586	10	AF191501	AF191501 Mus muscu
3	1110	82.8	2002	10	AB041596	AB041596 Mus muscu
4	1084.4	80.9	1363	10	AF107204	AF107204 Mus muscu
5	989.4	72.3	2000	9	AF229057	AF229057 Homo sapi
6	906.4	67.6	3348	9	AF109106	AF109106 Homo sapi
7	905.2	67.6	1513	9	AK001027	AK001027 Homo sapi
8	904.2	67.5	2279	9	AF107203	AF107203 Homo sapi
9	861.8	64.3	1475	9	AB060859	AB060859 Macaca fa
10	857.2	64.0	2372	6	AF134676	AF134676 Sequence
11	798.4	59.6	1547	9	AF094849	AF094849 Homo sapi
12	541.2	40.4	1623	9	HSB03056	AF387322 Mus muscu
13	295.2	22.0	1134	10	AF387322	AF387322 Mus muscu
14	295.2	22.0	1695	10	AF229055	AF229055 Mus muscu
15	285.4	21.3	115860	2	AC120661	AC120661 Rattus no
16	281.2	21.0	1721	10	BC027263	BC027263 Mus muscu
17	280	20.9	1876	9	HSP38A20	AL009266 H. sapien
18	277.6	20.7	1925	9	AF229058	AF229058 Homo sapi
19	276	20.6	1555	9	BC025281	BC025281 Homo sapi
20	275.6	20.6	1393	9	AY072786	AY072786 Homo sapi
21	274	20.4	1558	9	AK055213	AK055213 Homo sapi
22	195.8	14.6	959	10	AF229056	AF229056 Mus muscu
23	186	13.9	1538	9	BC013115	BC013115 Homo sapi
24	172	12.8	2971	9	HSFOX14	AF109120 Homo sapi
25	172	12.8	164538	9	AC005774	AC005774 Homo sapi
26	145.8	10.9	2457	3	AY118678	AY118678 Drosophil
27	133.8	10.0	169192	2	AC114672	AC114672 Mus muscu
28	133.8	10.0	235183	2	AC079427	AC079427 Mus muscu
29	123.8	9.2	629	9	HSFOX03	AF109109 Homo sapi
30	123.8	9.2	112976	9	AC027683	AC027683 Homo sapi
31	123.8	9.2	169765	9	AC006075	AC006075 Homo sapi
32	120.6	9.0	105907	2	AC026429	AC026429 Homo sapi
33	111	8.3	251	9	HSFOX09	AF109115 Homo sapi
34	109.6	8.2	1426	3	CEU14946	U14946 Caenorhabdi
35	94.2	7.0	219	9	HSFOX05	AF109111 Homo sapi
36	90.4	6.7	470	9	HSFOX02	AF109108 Homo sapi
37	89.2	6.7	203649	10	AL603843	AL603843 Mouse DNA
38	88.6	6.6	108130	2	AC098922	AC098922 Rattus no
39	88.6	6.6	156820	2	AC123391	AC123391 Rattus no
40	87.4	6.5	596	9	HSB323688	AJ323688 Homo sapi
41	87.4	6.5	617	9	HSB335811	AJ335811 Homo sapi
42	87.4	6.5	653	9	HSB335616	AJ335616 Homo sapi
43	87.4	6.5	675	9	HSB334013	AJ334013 Homo sapi
44	87	6.5	108315	9	HS41P2	AL049748 Human DNA
45	81.8	6.1	374	9	HSFOX08	AF109114 Homo sapi

ALIGNMENTS

RESULT 1	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
LOCUS	AX268800	Sequence 1 from Patent WO0174901.				
DEFINITION	AX268800					
ACCESSION	AX268800					
VERSION	AX268800.1	GI:16541860				
KEYWORDS						
SOURCE		Norway rat.				
ORGANISM		Rattus norvegicus				
		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
		Rattus.				
REFERENCE	1	Stanton,L.W. and White,R.T.				
AUTHORS		Secreted factors				
TITLE						